



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification⁶ : A01N 37/18, 43/04, C12Q 1/00, 1/02, 1/68, C12N 5/00, 5/06, 15/00, 15/06, 15/09, 15/10, 15/11, G01N 33/53</p>	A2	<p>(11) International Publication Number: WO 98/54963</p> <p>(43) International Publication Date: 10 December 1998 (10.12.98)</p>
<p>(21) International Application Number: PCT/US98/11422</p> <p>(22) International Filing Date: 4 June 1998 (04.06.98)</p> <p>(30) Priority Data: 60/048,915 6 June 1997 (06.06.97) US 60/048,882 6 June 1997 (06.06.97) US (Continued on the following page)</p> <p>(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown, MD 20874 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US).</p>		<p>Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mt. Nebo Road, Poolesville, MD 20857 (US). MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). FLORENCE, Charles [US/US]; (US). FLORENCE, Kimberly [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FAN, Ping [CN/US]; Apartment 302, 335 West Side Drive, Gaithersburg, MD 20878 (US). WEI, Ying-Fei [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [MM/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US).</p> <p>(74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p>
<p>(54) Title: 207 HUMAN SECRETED PROTEINS</p>		<p>Published With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.</p>

(Continued)

60/048.992	6 June 1997 (06.06.97)	US	60/057.631	5 September 1997 (05.09.97)	US
60/048.901	6 June 1997 (06.06.97)	US	60/057.769	5 September 1997 (05.09.97)	US
60/048.900	6 June 1997 (06.06.97)	US	60/057.643	5 September 1997 (05.09.97)	US
60/048.893	6 June 1997 (06.06.97)	US	60/057.645	5 September 1997 (05.09.97)	US
60/048.964	6 June 1997 (06.06.97)	US	60/057.668	5 September 1997 (05.09.97)	US
60/048.884	6 June 1997 (06.06.97)	US	60/057.635	5 September 1997 (05.09.97)	US
60/048.894	6 June 1997 (06.06.97)	US	60/057.627	5 September 1997 (05.09.97)	US
60/048.971	6 June 1997 (06.06.97)	US	60/057.567	5 September 1997 (05.09.97)	US
60/048.885	6 June 1997 (06.06.97)	US	60/057.666	5 September 1997 (05.09.97)	US
60/049.375	6 June 1997 (06.06.97)	US	60/057.764	5 September 1997 (05.09.97)	US
60/048.881	6 June 1997 (06.06.97)	US	60/057.644	5 September 1997 (05.09.97)	US
60/048.880	6 June 1997 (06.06.97)	US	60/057.765	5 September 1997 (05.09.97)	US
60/048.896	6 June 1997 (06.06.97)	US	60/057.762	5 September 1997 (05.09.97)	US
60/049.020	6 June 1997 (06.06.97)	US	60/057.775	5 September 1997 (05.09.97)	US
60/048.876	6 June 1997 (06.06.97)	US	60/057.634	5 September 1997 (05.09.97)	US
60/048.895	6 June 1997 (06.06.97)	US	60/057.777	5 September 1997 (05.09.97)	US
60/049.019	6 June 1997 (06.06.97)	US	60/057.628	5 September 1997 (05.09.97)	US
60/048.916	6 June 1997 (06.06.97)	US	60/057.776	5 September 1997 (05.09.97)	US
60/048.970	6 June 1997 (06.06.97)	US	60/057.760	5 September 1997 (05.09.97)	US
60/048.972	6 June 1997 (06.06.97)	US	60/057.761	5 September 1997 (05.09.97)	US
60/048.949	6 June 1997 (06.06.97)	US	60/057.771	5 September 1997 (05.09.97)	US
60/048.974	6 June 1997 (06.06.97)	US	60/057.770	5 September 1997 (05.09.97)	US
60/048.883	6 June 1997 (06.06.97)	US	60/057.649	5 September 1997 (05.09.97)	US
60/048.897	6 June 1997 (06.06.97)	US	60/057.774	5 September 1997 (05.09.97)	US
60/048.898	6 June 1997 (06.06.97)	US	60/057.648	5 September 1997 (05.09.97)	US
60/049.373	6 June 1997 (06.06.97)	US	60/057.642	5 September 1997 (05.09.97)	US
60/048.917	6 June 1997 (06.06.97)	US	60/057.629	5 September 1997 (05.09.97)	US
60/048.962	6 June 1997 (06.06.97)	US	60/057.778	5 September 1997 (05.09.97)	US
60/048.878	6 June 1997 (06.06.97)	US	60/057.763	5 September 1997 (05.09.97)	US
60/049.374	6 June 1997 (06.06.97)	US	60/057.584	5 September 1997 (05.09.97)	US
60/048.875	6 June 1997 (06.06.97)	US	60/057.654	5 September 1997 (05.09.97)	US
60/048.899	6 June 1997 (06.06.97)	US	60/057.646	5 September 1997 (05.09.97)	US
60/048.877	6 June 1997 (06.06.97)	US	60/057.662	5 September 1997 (05.09.97)	US
60/048.963	6 June 1997 (06.06.97)	US	60/057.650	5 September 1997 (05.09.97)	US
			60/057.661	5 September 1997 (05.09.97)	US
			60/057.647	5 September 1997 (05.09.97)	US
			60/070.923	18 December 1997 (18.12.97)	US

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

207 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,

5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°
10 C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even
25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include
30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes
35 as any 3' terminal

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5 The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be
10 single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability
15 or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

 The polypeptide of the present invention can be composed of amino acids joined
20 to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,
25 as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be
30 branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a
 nucleotide derivative, covalent attachment of a lipid or lipid derivative, cross-linking, cyclization, disulfide bond

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene is expressed primarily in melanocytes and, to a lesser extent, in testes, ovary, kidney and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer, disorders of neural crest derived cells including pigmentation defects, melanoma, reproductive organ defects, and defects of the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin,

reproductive, and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating disorders that arise from alterations in the number or fate of neural crest derived cells including cancers such as melanoma and defects of the developing reproductive system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in infant brain and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental disorders of the brain or lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating or diagnosing disorders associated with abnormal proliferation of cells in the Central nervous system and developing lung.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in breast lymph node and to a lesser extent in ovarian cancer and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune responses such as inflammation or immune surveillance for

tumors. This gene may be important for inflammatory responses associated with tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 236 as residues: Lys-45 to Val-50, Lys-69 to Arg-76.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune responses including those associated with tumor-induced inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in T-cells and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases involving T-cells such as inflammation, autoimmunity, and cancers including T-cell lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of T-cells and other cells of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating T-cell based disorders such as inflammatory diseases, autoimmune disease and tumors including T-cell lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation, autoimmunity, infection, or disorders involving activation of monocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 238 as residues: Asp-19 to Arg-31.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing or treating diseases that result in activation of monocytes including infections, inflammatory responses or autoimmune diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene shares sequence homology with terminal deoxynucleotidyltransferase which is thought to be important in catalyzing the elongation of oligo- or polydeoxynucleotide chains.

This gene is expressed primarily in activated human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly those of the blood such as leukemia and deficiencies in neutrophils such as neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to terminal deoxynucleotidyltransferase indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and differential diagnosis of acute leukemia's. Alternatively, this gene may function in the proliferation of neutrophils and be useful as a treatment for neutropenia, for example, following neutropenia as a result of chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The contig exhibits a reasonable homology to the human chorionic gonadotropic (HCG) analogue-GT beta-subunit as disclosed in U.S. Patent No. 5,508,261 and PCT Publication No. WO 92/22568. There is a high degree of conservation of the structurally important cysteine residues in these identities.

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene is expressed primarily in IL-1- and LPS-induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 241 as residues: Ser-14 to Pro-22, Leu-43 to Val-53.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system, including inflammatory diseases and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 242 as residues: Tyr-22 to His-35.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the immune system since expression is primarily in neutrophils, and may be useful as a growth

factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia following chemotherapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune dysfunctions including cancer of the T lymphocytes and autoimmune disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune disorders particularly of T-cell origin and may act as a growth factor for particular subsets of T-cells such as CD4 positive cells which would make this a useful therapeutic for the treatment of HIV and other immune compromising illnesses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in fetal tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor or differentiation factor for particular cell types in the developing fetus and may be useful in replacement or other types of therapy in cases where the gene is expressed aberrantly.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in T-cells and to a lesser extent in tumor tissue including glioblastoma, meningioma, and Wilm's tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the immune system including autoimmune conditions such as rheumatoid arthritis, inflammatory disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 245 as residues: Thr-9 to Ser-14.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis/ modulation of immune function disorders, including rheumatoid arthritis and inflammatory responses.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed primarily in placenta and to a lesser extent in fetal liver and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the hematological and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells in the treatment of chemotherapy patients or kidney disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed primarily in stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hematopoietic disorders including cancer, neutropenia, anemia, and thrombocytopenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells, in particular following chemotherapy treatment.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene shares sequence homology with epsilon-COP from *Bos taurus* which is thought to be important as a component of coatamer, a complex of seven proteins, that is the major component of the non-clathrin membrane coat. Preferred polypeptides encoded by this gene comprise the following amino acid sequences:

MAPPAPGPASGGSGEVDELFDVKNAFYIGSYQQCINEAXXVKLSSPERDVERD

VFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMFADYLAHESRRDSIVAELDRE
 MSRSXDVTNTTFLMLAASIYLHDQNPDAALRALHQGDSLECTAMTVQILLKLD
 RLDLARKELKRMQDLDEDATLTQLATAWVSLATGGEKLQDAYYTFQEMADKCS
 PTLNLLNGQAACHMAQGRWEAAEGLLQEALDKDSGYPETLVNLIVLSQHLGKP
 5 PEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRLVLQYAPSAEAGPELSGP
 (SEQ ID NO:458); or RDVERDVFLYRAYLAQRKFGVVLDEIKPSSAPELQAVRMF
 ADYLAHESRRDSIVAELDREMSRSXDVTNTTFLMLAASIYLHDQNPDAALRALH
 QGDSLECTAMTVQILLKLDRLDLARKELKRMQDLDEDATLTQLATAWVSLATG
 GEKLQDAYYTFQEMADKCSPTLNLNNGQAACHMAQGRWEAAEGLLQEALDKD
 10 SGYPETLVNLIVLSQHLGKPPEVTNRYLSQLKDAHRSHPIKEYQAKENDFDRL
 VLQYAPSA (SEQ ID NO:459).

This gene is expressed primarily in activated monocytes and T-cells, and to a lesser extent in multiple other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, immunomodulation, specifically relating to transport problems in these
 cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
 providing immunological probes for differential identification of the tissue(s) or cell
 20 type(s). For a number of disorders of the above tissues or cells, particularly of the
 immune, expression of this gene at significantly higher or lower levels may be routinely
 detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
 taken from an individual having such a disorder, relative to the standard gene
 25 expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution and homology to epsilon-COP indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for treating
 /diagnosing problems with the cellular transport of proteins that may result in
 30 immunologic dysfunction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with an RNA
 helicase which is thought to be important in polynucleotide metabolism. The translation
 35 product of this contig exhibits good homology to the LbeIF4A antigen of *Leishmania*
braziliensis. The LbeIF4A antigen, or immunogenic portions of it, can be used to
 induce protective immunity against leishmaniasis, specifically *L. donovani*, *L. chagasi*.

L. infantum, *L. major*, *L. braziliensis*, *L. panamensis*, *L. tropica* and *L. guyanensis*. It can also be used diagnostically to detect *Leishmania* infection or to stimulate a cellular and/or humoral immune response or to stimulate the production of interleukin-12.

This gene is expressed primarily in colon cancer and to a lesser extent in
5 pituitary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers particularly of the colon. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 249 as residues: Glu-93 to Ala-98, Gln-150 to Leu-156, Leu-220 to Leu-231, Leu-268 to Arg-273, Val-324 to Pro-341, Arg-372 to Asn-380, Ser-405 to Gly-410, Phe-426 to Ala-433, Glu-458 to Asp-470, Arg-506 to Ser-547.
20

The tissue distribution and homology to RNA helicase indicates that polynucleotides and polypeptides corresponding to this gene are useful for development of diagnostic tests for colon cancer.
25

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

The translation product of this contig has sequence homology to a cytoplasmic protein that binds specifically to JNK designated the JNK interacting protein-1 or JIP-1 in mice. JIP-1 caused cytoplasmic retention of JNK and inhibition of JNK-regulated
30 gene expression.

This gene is expressed primarily in brain including pituitary cerebellum frontal cortex, fetal brain and to a lesser extent in the kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of the central nervous system disorders including ischemia, epilepsy, Parkinson's disease, and schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, the translation product of this contig may suppress the effects of the JNK signaling pathway on cellular proliferation, including transformation by the Bcr-Abl oncogene. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 250 as residues: Pro-6 to Ser-26, Ala-30 to Asp-41, Gly-55 to Ser-61, Gly-74 to Thr-80, Tyr-117 to Ala-123, Tyr-167 to Asp-172, Ala-212 to Cys-223, Pro-239 to Tyr-244.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for enhanced survival and/or differentiation of neurons as a treatment for neurodegenerative disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The translation product of this gene shares sequence homology with a liver stage antigen from a protozoan parasite.

This gene is expressed primarily in fetal tissue and to a lesser extent in activated T-cells and other immune cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and diseases of immune function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to a protozoan antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/immune modulation of parasitic infections.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 19

Preferred polypeptide encoded by this gene comprise the following polypeptide sequences:

MKAIGIEPSLATYHHIIRLFDQPGDPLKRSSFIIYDIMNELMGKRFSPKD
 PDDDKFFQSAMSICSSLRDLELAYQVHGLLKTGDNWKFIGPDQHRNFYYSKFF
 10 DLICLMEQIDVTLKWYEDLIPSA YFPHSQTMHLLQALDVANRLEVIPKIWER
 (SEQ ID NO:460); and/or KDSKEYGHTFRSDLREEILMLMARDKHPPPELQVAF
 ADCAADIKSAYESQPIRQTAQDWPATSLNCIALFLRAGRTQEAWKMLGLFRKH
 NKIPRSELLNELMDSAKVSNSPSQAIEVVELASAFSLPICEGLTQRVMSDFAINQ
 EQKEALSNLTALTSDSDTDSSSDSDSDTSEGK (SEQ ID NO:461). Polynucleotides

15 encoding such polypeptides are also provided.

This gene is expressed primarily in stromal and CD34 depleted bone marrow cells and to a lesser extent in tissues of embryonic origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
 20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of hematologic origin including cancers and immune dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
 25 the hematopoietic and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 30 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 252 as residues: Ser-28 to Gln-34.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a growth factor for hematopoietic stem cells or progenitor cells which may be useful in the treatment of chemotherapy patients
 35 suffering from neutropenia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

Preferred polypeptide fragments can be found in an alternative open reading frame. These preferred polypeptides comprise the amino acid sequence:

5 MSSDNESDIEDLDKLELRRLRDKHLKEIQDLQSRQKHEIESLYTKLGKVPPAVI
 IPPAAPLSGRRRRPTKSKGSKSSRSSSLGNKSPQLSGNLSGQSAASVLHPQOTL
 HPPGNIPESGQNQLLQPLKPSRSSDNL YSAFTSDGAISVPSLSAPGQGTSSSTNTV
 GATVNSQAAQAQPPAMTSSRKGTFTDDLHKLVDNWARDAMNLSGRRGSKGH
 MNYEGPGMARKFSAPGQLCISMTSNLGGAPISA.AAS.ATSLGHFTKSMCPPQQY
 GFPATPFGAQWSGTGGPAPQPLGQFQPVGTASLQNFNISNLQKSISNPPGSNL
 10 RTT (SEQ ID NO:462); IQDLQSRQKHEIESLYTKLGKVPPAVIIPPAAPLSGRRRR
 PTKSKGSKSSRSSSLGNKSPQLSGNLSGQSAASVLHPQOTLHPPGNIPESGQN
 QLLQPLKPSRSSDNL YSAFTSDGAISVPSLSAPGQGTSSST (SEQ ID NO:463);
 TSDGAISVPSLSAPGQGTSSSTNTVGATVNSQAAQAQPPAMTSSRKGTFTDDLH
 (SEQ ID NO:464); KGHMNYEGPGMARKFSAPGQLCISMTSNLGGAPISAAS
 15 ATSLGHFTK (SEQ ID NO:465); QPLKPSRSSDNL YSAFTSDGAISVPSLSAPG
 (SEQ ID NO:466). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments.

This gene is expressed in fetal liver and tissues associated with the CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as
 20 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, liver and CNS diseases. Similarly, polypeptides and antibodies directed
 to these polypeptides are useful in providing immunological probes for differential
 identification of the tissue(s) or cell type(s). For a number of disorders of the above
 25 tissues or cells, particularly of the liver and CNS, expression of this gene at
 significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 30 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
 epitopes include those comprising a sequence shown in SEQ ID NO: 253 as residues:
 Gln-26 to Lys-34.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for diagnosis and treatment for liver diseases such
 35 as hepatocellular carcinomas and diseases of the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

In an alternative reading frame, this gene shows sequence homology to two recently cloned genes, karyopherin beta 3 and Ran_GTP binding protein 5. (See Accession Nos. gil2102696 and gnlIPIDle328731.) The Ran_GTP binding protein is related to importin-beta, the key mediator of nuclear localization signal (NLS)-dependent nuclear transport. Based on homology, it is likely that this gene may activity similar to the RAN_GTP binding protein. Preferred polypeptide fragments comprise the amino acid sequence: VRVAAESMXLLLECAXVRGPEYLTQMWHFMCDALIKA IGTEPDSDVLSEIMHSFAK (SEQ ID NO:467). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in thymus tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

This gene is expressed primarily in prostate and osteoclastoma tissues. Preferred polypeptide fragments also comprise the amino acid sequence: MEINNQNCFIVIDLVRTVMENGVEGLLIFGAFLPESWLIGVRCSSSEPPKALLLIL AHSQKRRLDGWSFIRHLRVHYCVSLTIHFS (SEQ ID NO:468). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone and prostate diseases, and cancers, particularly of the bone and prostate. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone and prostate systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 255 as residues: Met-1 to Ser-11.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for bone and prostate disorders, especially cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 23

This gene shares sequence homology with the FK506-binding protein (FKBP-13) family, a known cytosolic receptor for the immunosuppressants. Recently, another group has cloned a very similar gene, recognizing the homology to FK506-binding protein family, calling their gene FKBP23. (See Accession No. 2827255.)

This gene is expressed primarily in lymphoid tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample, especially for those susceptible to immune suppressant therapies and for diagnosis of diseases and conditions, which include, but are not limited to, immune suppressant disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 256 as residues: Ala-19 to Val-31, Arg-38 to Gly-49, Ala-61 to Lys-66, Tyr-68 to Pro-78, Gly-116 to Ala-121, Asp-154 to Ser-162, Glu-173 to Gln-186, Phe-194 to Gly-203. Pro-207 to Val-212.

The tissue distribution and homology to FKBP-12 and -13 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for immune suppressant disorders.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in the brain and in the retina. This gene maps to chromosome 8, and therefore can be used in linkage analysis as a marker for chromosome 8.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and ocular associated disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of
15 disorders of the above tissues or cells, particularly of the disorders of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 257 as residues: Cys-34 to Asp-40.

The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders
25 including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans, and retinoblastoma. Expression in the brain indicates a role in the is useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive
30 disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

This gene shows sequence homology to a newly identified class of proteins expressed in the nervous system, called stathmin family. (See Accession No. 2585991;
35 see also Eur. J. Biochem. 248 (3), 794-806 (1997).) The stathmin family appears to be an ubiquitous phosphoprotein involved as a relay integrating various intracellular signaling pathways. These pathways affect cell proliferation and differentiation.

Preferred polypeptide fragments comprise the amino acid sequence:

QDKHAEVRKNKELKEEASR (SEQ ID NO:469); QQDLSPWAAPVGCPLXXASX
TCHXLPLSGCLRRQSXSLPVVAXLCFWFSCPLASLFVPGQPCVTCPFPSLPFQD
KHAEVRKNKELKEEASR (SEQ ID NO:470). Also preferred are the

5 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
10 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
20 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 26

The polynucleotide sequence of this gene contains a domain similar to a Flt3 ligand peptide. Preferred polypeptide fragments comprise the amino acid sequence: PTRCCTTQPCRSSARRPCWVPMVPSPEGREXQPTCPS (SEQ ID NO:471). Thus,
30 this gene may have activity as binding to Flt3 receptors, a process known to promote angiogenesis and/or lymphangiogenesis.

This gene is expressed in human tonsil, and to a lesser extent in teratocarcinoma, placenta, colon carcinoma, and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample
35 and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the tonsil, as well as cancers, such as colon, reproductive, and kidney cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful

in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tonsils, colon, reproductive organs, and kidneys, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 259 as residues: Pro-22 to Glu-33.

The tissue distribution in tonsil and several cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the tonsil or colon, such as tonsillitis, inflammatory diseases involving nose and paranasal sinuses, especially during the infection of influenza, adenoviruses, parainfluenza, rhinoviruses. The gene may also be useful in the diagnosis and treatment of neoplasms of nasopharynx or colon origins.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In an alternative reading frame exists a large open reading frame that encodes a preferred polypeptide. Preferred polypeptide fragments comprise the amino acid sequence:

MKRSLNENSARSTAGCLPVPLFNQKKRNRQPLTSNPLKDDSGISTPSDNYDFP
PLPTDWAWAEAVNPEXAPVMKTVDTGQIPHVS SRPLRSQDSVFNSIQSNTGRSQ
GGWSYRDGNKNTSLKTWXXKNDFKPQCKRTNLVANDGKNSCPMSSGAQQQK
QLRTPEPPNLSRNKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNFQQNQY
KXQMLDDIPEDNTLKETS LYQLQFKEKASSLR IISAVIESMKYWREHAQKT VLL
FEVLAVLDSAVTPGPYYSKTFLMRDGKNTLPCVFYEIDREL PRLIRGRVHRCVG
NYDQKKNIFQCVSVRPASVSEQKTFQAFVKIADVEMQYYINVMNET (SEQ ID
NO:472); SQDSVFNSIQSNTGRSQGGWSYRDGNKNTSLKTWXXKNDFKPQCKR
(SEQ ID NO:473); NKETELLRQTHSSKISGCTMRGLDKNSALQTLKPNF (SEQ ID
NO:474); SSLRIISAVIESMKYWREHAQKT VLLFEVLAVLDSAVTPGPYYSKTFLM
(SEQ ID NO:475); and PRLIRGRVHRCVGN YDQKKNIFQCVSVRPASVSEQT
FQAFV (SEQ ID NO:476).

This gene is expressed primarily in human testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

related to, male reproductive disorders, including cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a hormone with reproductive or other systemic functions; contraceptive development; male infertility of testicular causes, such as Klinefelter's syndrome, varicocele, orchitis; male sexual dysfunctions; testicular neoplasms; and inflammatory disorders such as epididymitis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases relating to T cells, as well as cancer in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders. Moreover, since the gene was isolated from an apoptotic cell and based on the understanding of the relationship of apoptosis and cancer, it is likely that this gene may play a role in the genesis of cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, gastrointestinal disorders. Similarly, polypeptides and antibodies directed
to these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the gastrointestinal system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the diagnosis and treatment of gastrointestinal
diseases.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

The translation product of this gene shares sequence homology with C44C1.2
gene product of *Caenorhabditis elegans* with unknown function. Preferred polypeptide
fragments comprise the amino acid sequence:

GVFRPCVCGRPASLTCSPLDPEVGPYCDTPTMRTL FNLLWLALACSPVHTTLSK
25 SDAKKAASKTLEKSQFSDKPVQDRGLVVTDLKAESVVLEHRSYCSAKARDRH
FAGDVLGYVTPWNSHGYDVTKVFGSKFTQISPVWLQLKRRGREMF EVTGLHD
VDQGWMRAVRKHAKGLHIVPRLLFEDWTYDDFRNVLDSEDEIEELSKTVVQVA
KNQHFDGFFVEVWNQLLSQKR VGLIHMLTHLAEALH QARLLALLVIPPAITPGT
DQLGMFTHKEFEQLAPVLDGFS LMTYDYSTAHQPGPNAPLSWVRACVQVLDP
30 KXKWRTKSSWGSTSMXWTXRPXDARXPVVGXRXIQXLKDHXPRMVLD SK
PQ (SEQ ID NO:477); TCSPLDPEVGPYCDTPTMRTL FNLLWLALACSPVHTTLS
(SEQ ID NO:478); LVVTDLKAESVVLEHRSYCSAKARDRH FAGDVLGYVTPW
NSHGYDVTKVFGSKF (SEQ ID NO:479); REMFEVTGLHDVDQGWMRAVRK
HAKGLHIVPRLLFEDWTYDDFRNVLDSEDE (SEQ ID NO:480); HFDGFFVEVW
35 NQLLSQKR VGLIHMLTHLAEALH QARLLALLVIPPAITPGTDQLGM (SEQ ID
NO:481); DGFSLMTYDYSTAHQPGPNAPLSWVRACVQVLDPKXKWRTKSSW
GST (SEQ ID NO:482). Also preferred are polynucleotide fragments encoding these

polypeptide fragments. This gene maps to human chromosome 11, and therefore is useful in linkage analysis as a marker for chromosome 11.

This gene is expressed primarily in human T cells and to a lesser extent in human colon carcinoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 263 as residues: Leu-21 to Ala-30, Ser-38 to Asp-47, Pro-87 to Asp-94, Leu-197 to Thr-204, Pro-256 to Ser-262, Thr-277 to Arg-282, Thr-293 to Trp-303.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders and gastrointestinal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The translation product of this gene shares sequence homology with Ribosomal protein L11 of *Caenorhabditis elegans*. (See Accession No. 156201.) Preferred polypeptide fragments comprise the amino acid sequence:

ERGVSINQFCKEFNERTKDIKEGIPLPTKILVKPDRTFEIKIGQPTVSYFLKAAAG
IEKGARQTGKEVAGLVTLKHVYEIARIKAQDEAFALQDVPLSSVVRSIIGSARSL
GIRVVKDLSSEELAAFQKERAIFLAAQKEADLAAQEEAAKK (SEQ ID NO:483).

Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in human embryo tissue and to a lesser extent in human epithelioid sarcoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development disorders and epithelial cell cancer. Similarly, polypeptides and antibodies

directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic and epithelial cell systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 264 as residues: Lys-34 to Gly-40.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental disorders and epithelial cancer.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in resting T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders of immune system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is believed to reside on chromosome 1. Accordingly, polynucleotides derived from this gene are useful in linkage analysis as chromosome 1 markers.

This gene is expressed primarily in prostate and to a lesser extent in soares adult brain, human umbilical vein endothelial cells, and amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the urinary system and nervous system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and treatment of disorders of the urinary and nervous systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

This gene shares sequence homology with R05G6.4 gene product. (See Accession No. gill326338.) This gene also shares sequence homology with the cyclophilin-like protein Cyp-60. (See Accession No. 1199598, see also Biochem. J. 314 (1), 313-319 (1996).) Preferred polypeptide fragments comprise the amino acid sequence:

AVYTYHEKKKDTAASGYGTQNI RLSRDAVKDFDCCCLSLQPCHDPVVTDPDGYL
YERE AILEYILHQKKEIARQMKA YEKQRGTRREEQKELQRAASQDHVRGFLEKE
SAIVSRP LNPFTAKALSGTSPDDVQPGPSVGPPSKDKDKVLPSFWIPSLTPEAK
ATKLEKPSRTVTCMSGKPLRMSDLTPVHFTPLDSSVDRVGLITRSEYVCAVT
RDSLSNATPCAVLRPSGAVVTLECEVKLIRKDMVDPVTGDKLTDRDIIVLQRGT
(SEQ ID NO:484); YLYERE AILEYILHQKKEIARQMKA YEKQRGTRREEQKELQ
RAASQDHVRGFLE (SEQ ID NO:485); and FTAALSGTSPDDVQPGPSVGPP
SKDKDKVLPSFWIPSLTPEAKATKLEKPSRTVTCMSGKPL (SEQ ID NO:486).

Also preferred are polynucleotide fragments that encode these polypeptide fragments.

This gene is expressed primarily in human testis and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders and in particular testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system. Expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the male reproductive system and in particular of testicular cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

The translation product of this gene shares sequence homology with Lpe5p of *Saccharomyces cerevisiae* which is thought to be important in the metabolism of phospholipids.

This gene is expressed primarily in liver and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and nervous systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 268 as residues: Pro-14 to Leu-20, Lys-28 to Asn-38, Arg-109 to Arg-114, Lys-119 to Asn-124, Glu-152 to Leu-157, Pro-172 to Val-180.

The tissue distribution and homology to Lpe5p of *Saccharomyces cerevisiae* indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of metabolic and nervous disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

This gene shares sequence homology with the nuclear ribonucleoprotein U (HNRNP U), encoded by *C. elegans* (See Accession gill703576.) Preferred polypeptide fragments comprise the amino acid sequence:

5 MDTSENRPENDVPEPPMPIADQVSNDDEEGSVDEEEKKESLPSFKRKISVV
SATKGVPAAGNSDTEGGQPGRKRRWGASTATTQKKPSISITTESLKS LIPDIKPL
AGQEAVVDLHADDSETERNGDDGTHDKGLKICRTVTQVVP AEGQENGQ
REEEEEKEPEAEPPVPPQVSVEVALPPPAEHEVKKVTLGDTLTRRSISQQKSGV
SITIDDPVRTAQVPSPPRGKISNIVHISNLVRPFTLGQLKELLGRTGTLVEEAFWI
10 DKIKSHCFVTYSTVEEAVATRTALHGVKWPQSNPKFLCADYAEQDEL DYHRGL
LVDRPSETKTEEQGIPRPLHPPPPPVQPPQHPRAEQREQERAVREQWAERERE
MERRERTRSEREWDRDKVREGPRSRSRSRXRRRKERAKSKEKKSEKKEKAQE
EPPAKLLDDLFRKTKAAPCIYWLPLTDSQIVQKEAERAERAKEREKRRKEQEEE
EQKEREKEAERERNRQLEREKRREHSRERDRERERERERDRGDRDRDRERDRE
15 RGRERDRRDTKRHSRSRSTPVRDRGGR (SEQ ID NO:488). Also preferred are
the polynucleotide fragments encoding this polypeptide fragments.

This gene is expressed primarily in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the male reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of
25 this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
30 disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of male reproductive disorders.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory diseases and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the amygdala, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases and reproductive disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene shares sequence homology with human opsonin protein P35 fragment. (See Accession No. R94181.) The opsonin protein activates the phagocytosis of pathogenic microbes by phagocytic cells. Preferred polypeptide fragments comprise the amino acid sequence: GCDSCPPHLPREAFAQDTQAEGECSSRAERADMCPDAP PSQEVPEGPGAAP (SEQ ID NO:489). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed in immune-related tissues such as thymus, macrophage, T cells and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and infectious disease, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 271 as residues: Lys-9 to Arg-14, Met-38 to Asp-51.

The tissue distribution indicates that polynucleotides and polypeptides
 5 corresponding to this gene are useful for diagnosis and treatment of immune disorders, as well as the treatment and/or diagnosis of infectious disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 39

The translation product of this gene shares sequence homology with alpha-2
 10 type I collagen which is thought to be important in tissue repair. (See, e.g., 211607.) Preferred polypeptide fragments comprise the amino acid sequence: PQLPSCGRPWP
 PGTASVFQSHTQGPREDPDPCRAQGSAGTHCPISLSPPRQ (SEQ ID NO:490).
 Also preferred are the polynucleotide sequences encoding these polypeptide sequences.

This gene is expressed primarily in the brain and to a lesser extent in the kidney
 15 and thymus

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, brain, kidney, and immune disorders. Similarly, polypeptides and
 20 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the brain, kidney, and immune disorders,
 expression of this gene at significantly higher or lower levels may be routinely detected
 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
 25 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
 an individual having such a disorder, relative to the standard gene expression level, i.e.,
 the expression level in healthy tissue or bodily fluid from an individual not having the
 disorder.

The tissue distribution and homology to alpha-2 type I collagen indicates that
 30 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
 and treatment of tissue repair, and brain, kidney, immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

The translation product of this gene shares sequence homology with mini-
 35 collagen which is thought to be important in tissue repair tumor metastasis. (See
 Accession No. gnllPID1006976.) Preferred polypeptide fragments comprise the
 amino acid sequence: PGFRGPSGLGCSFFPRSLGRVLPQGCQRPGAHAD

SSPPPTP (SEQ ID NO:491). Also preferred are polynucleotides encoding this polypeptide fragment.

This gene is expressed in ovarian cancer and to a lesser extent in dendritic cells and smooth muscle.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumor metastasis and tissue repair. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor metastasis and tissue repair, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
15 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 273 as residues: Asn-2 to His-11.

The tissue distribution and homology to mini-collagen gene indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumor metastasis and tissue repair.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

This gene shares sequence homology with the HIV TAT protein. (See
25 Accession No. 328416.) Preferred polypeptide fragments comprise the amino acid sequence: EDLKKPDPASLRAASC GEGKKRKACKNCTCGLAE E E L E K E K
SREQMSSQPKSACGNCYLGD AFR CASC PYLGMPAFKPG EKVLLS (SEQ ID NO:492); EDLKKPDPASLRAASC GEGKKRKACKNCTCGLAE E E L E K E K
SREQMSSQPKSACGNCYLGD AFR CASC PYLGMPAFKPG EKVLLS DSNLHD
30 (SEQ ID NO:493); CGNCYLGD AFR CASC PYLGMPAFKPG EKVLLS D S
(SEQ ID NO:494); SCGEGKKRKACKNCTCGLAE E E L E K E (SEQ ID NO:495);
SQPKSAC GNCYLGD AFR CASC (SEQ ID NO:496); and REAGQNSERQYVS
LSRD (SEQ ID NO:497). Also preferred are polynucleotide fragments encoding these
polypeptide fragments.

35 This gene is expressed primarily in the infant brain and to a lesser extent in the breast and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain, testes and breast disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, testes and breast disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 274 as residues: Pro-7 to Val-15.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of brain, testes and breast, and other related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

This gene is expressed primarily in the infant brain, human cerebellum, and to a lesser extent in medulloblastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, brain related disorders and medulloblastoma and other brain cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain related disorders and brain cancers, including medulloblastoma, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 275 as residues: Thr-41 to Glu-47.

FEATURES OF PROTEIN ENCODED BY GENE NO: 44

This gene is expressed primarily in the fetal brain, cerebellum and to a lesser extent in the placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal cell related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell related disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 277 as residues: Thr-20 to Gly-28.

The tissue distribution and homology to proline-rich protein genes indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with precerebellin of human, which is thought to be important in synaptic physiology. (See Accession No. gi1180251.) It has been observed that cerebellin-like immunoreactivity is associated with Purkinje cell postsynaptic structures. Thus, it is likely that this gene also have synaptic activity. Preferred polypeptide fragments comprise the amino acid sequence: QEGSEPVLLGECLVVCEPGRAAAGGPGGAALGEAPPGRVAFXAV RSHHHEPAGETGNGTSGAIYFDQVLVNEGGGFDRASGSFVAPVRGVYSFRFH VVKVYNRQTVQVSLMLNTWPVISAFANDPDVTREAAATSSVLLPLDPGDRVSLR LRRGXSTGW (SEQ ID NO:499). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in cerebellum and infant brain. By Northern analysis, a single transcript of 2.4 kb was observed in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain related disorders, brain cancers, and medulloblastoma.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

The translation product of this gene shares sequence homology with a phosphotyrosine-independent ligand for the lck SH2 domain which is thought to be important in signal transduction related to phosphotyrosine-independent ligand for the lck SH2 domain. (See Accession No. gill184951.) Preferred polypeptide fragments
10 comprise the amino acid sequence: ESSGQARTLADPGPGWPRQQGMCFGSLT
GLSTTPHGFLT VSAEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAAL
DTIQYSKH (SEQ ID NO:498). Also preferred are polynucleotide fragments encoding this polypeptide fragment. It is likely that this gene is a new member of a family of phosphotyrosine-independent ligands for the lck SH2 domains.

15 This gene is expressed primarily in the placenta and to a lesser extent in endothelial cells and neutrophil.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
20 not limited to, reproductive, cardiovascular, immune, and infectious diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, reproductive, and immune system, and infectious diseases, expression
25 of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
30 disorder.

The tissue distribution and homology to a phosphotyrosine-independent ligand for the lck SH2 domain indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cardiovascular, reproductive, and immune system diseases, as well as infectious diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene shares sequence homology with a 12 kD nucleic acid binding protein of Feline calicivirus which is thought to be important in viral replication. (See Accession No. 59264)

- 5 This gene is expressed primarily in human cardiomyopathy and to a lesser extent in T helper cells, fetal brain and synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiomyopathy as well as viral infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 280 as residues: Trp-20 to Cys-26.

- 20 The tissue distribution in cardiomyopathy and homology to viral 12 kD nucleic acid binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of cardiomyopathy, including those caused by ischemic, hypertensive, congenital, valvular, or pericardial abnormalities.
- 25 The gene expression pattern may be the consequence or the cause for these conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

- 30 The translation product of this gene shares sequence homology with tumor necrosis factor related gene product which is thought to be important in tumor necrosis, bacterial and viral infection, immune diseases and immunoreactions.

This gene is expressed primarily in colon and to a lesser extent in ovarian and breast cancers.

- 35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary or breast origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

not limited to, neuronal cell signal transduction and synaptic physiology. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal cell signal transduction and synaptic physiology expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene or gene family indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal cell related disorders.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

This gene is expressed in fetal liver and spleen, and to a lesser extent in bone marrow, umbilical vein, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the immune system, particularly hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 279 as residues: Asp-30 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Tumor necrosis factors indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of cancers of colon, ovary and breast origins, because TNF family members are known to be involved in the tumor development.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with mucins, such as epithelial mucin, which is thought to be important in extracellular matrix functions such as protection, lubrication and cell adhesion (See for example Accession No. R68002). Preferred polypeptide fragments comprise the following amino acid sequence: PRSRPALRPGRQRPPSHSATSGVLRPRKKPDP (SEQ ID NO:500).

Also preferred are polynucleotide fragments encoding these polypeptide fragments. Moreover, this gene maps to chromosome 22q11.2-qter, and therefore, can be used as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in corpus colosum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors, especially of corpus colosum, as well as metastatic lesions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the corpus colosum and other solid tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to mucins indicates that polynucleotides and polypeptides corresponding to this gene are useful for serum tumor markers or immunotherapy targets because tumor cells have greatly elevated level of mucin expression and shed the molecules into the epithelial tissues.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in CD34 depleted buffy coat cord blood and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disorders and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
15 a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34 depleted buffy coat cord blood and primary dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.
25 Secreted or cell surface proteins in the above tissue distribution often are involved in cell activation (e.g. cytokines) or molecules involved in cell surface activation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 51

The translation product of this gene shares sequence homology with Interferon
30 induced 1-8 gene encoded polypeptide which is thought to be important in binding to retroviral rev responsive element. Preferred polypeptide fragment comprise the following amino acid sequences: MTLITPSXKLTFXKGKNSWSSRACSSSTLVDP (SEQ ID NO:501). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

35 This gene is expressed primarily in CD34 positive cells and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, retroviral infection, such as AIDS, and other immune disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 284 as residues: Gln-51 to Trp-62.

The tissue distribution and homology to interferon induced gene 1-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of retroviral infection including HIV. The factor may be involved in viral stability or viral entry into the cells. Alternatively, the virus/factor complex may elicit the cellular immune reaction.

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene shares sequence homology to immunoglobulin lambda chain (See Accession No. 2865484). Therefore it is likely that this gene has activity similar to an immunoglobulin lambda chain. Preferred polypeptide fragments comprise the following amino acid sequence: GHPSALSIAPSDGSQLPCDEVYPYGEAHVTRYCKKPLTNS HLETEAQSSSL (SEQ ID NO:502). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, Hodgkin's lymphoma and other immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 285 as residues: Pro-27 to Thr-32.

5 The tissue distribution in Hodgkin's lymphoma and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune
10 functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

This gene has extensive homology to cDNA for Homo sapiens mRNA for the
15 ISLR gene(See Accession No. AB003184). This protein is considered to be a new member of the Ig superfamily and contains a leucine-rich repeat (LRR) with conserved flanking sequences and a C2-type immunoglobulin (Ig)-like domain. These domains are important for protein-protein interaction or cell adhesion, and therefore it is possible that the novel protein ISLR may also interact with other proteins or cells. The ISLR gene
20 was mapped on human chromosome 15q23-q24 by fluorescence in situ hybridization (See Medline Article No. 97468140). Homology to the ISLR gene has been confirmed by another independent group as well (See Accession No. Hs.102171)

This gene is expressed in a number of tissues including human retina, heart, skeletal muscle, prostate, ovary, small intestine, thyroid, adrenal cortex, testis,
25 stomach, spinal cord, fetal lung and fetal kidney tissues, colon, tonsil and stomach cancer, and to a lesser extent in endometrial stromal cells treated with estradiol, breast tissue, synovium, lymphoma, and number of other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of colon, ovary and breast origins. However, due to the wide range of expression in various tissues, protein may play a vital role in the development of cancer in other tissues as well, not just those mentioned above. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the colon, ovary and breast, expression of this gene at significantly higher or lower levels may be routinely

detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, this gene maps to chromosome 15q23-q24, and therefore, can be used as a marker in linkage analysis for chromosome 15.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

Gene has homology to multidrug resistance gene 1 (See Accession No. P06795). Preferred polynucleotide fragments comprise the following sequence:
GCTTCGTGTCCAACCCTCTTGCCCTTCGCCTGTGTGCCTGGAGCCAGTCCCA
CCACGCTCGCGTTTCCTCCTGTAGTGCTCACAGGTCCCAGCACCGATGGCA
TTCCCTTTGCCCTGAGTCTGCAGCGGGTCCCTTTTGTGCTTCCTTCCCCTCA
GGTAGCCTCTCTCCCCCTGGGCCACTCCCGGGGGTGAGGGGGTTACCCCTT
CCCAGTGTTTTTTATTCCTGTGGGGCTCACCCCAAAGTATTAAAAGTAGCTTT
GTAA (SEQ ID NO:503). Also preferred are polypeptide fragments encoded by these polynucleotide fragments.

This gene is expressed primarily in lung, esophagus, leukemia (Jurkat cells) and breast cancers and to a lesser extent in macrophages treated with GM-CSF fetal tissues and wide range of tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer of wide range of origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the solid tumors, lung and leukemia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Furthermore, due to the high expression level in lung tissue and the proposed function of the multidrug resistance protein 1 gene as the efflux pump responsible for low-drug accumulation in multidrug-resistant cells, protein as well mutants thereof, may also be beneficial as a target for gene therapy, particularly for the chronic patient. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 287 as residues: Met-1 to Lys-16.

The tissue distribution in wide range of cancers and fetal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of cells in active proliferation, such as cancers. The gene products may be used for cancer markers or immunotherapy target.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene maps to the X chromosome.

This gene is expressed primarily in the brain and to a lesser extent in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders, including sex-linked disorders, of the above tissues or cells, particularly of the neurological, developmental systems, and cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, this gene maps to the X chromosome, and therefore, may be used as a marker in linkage analysis for this chromosome.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Klinefelter's, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

5 The translation product of this gene shares sequence homology with paxillin which is thought to be important in mediating signal transduction from growth factor receptors to the cytoskeleton. Preferred polynucleotide fragments comprise the following sequence: TGGCTCACTGTCTTACAATCACTGCTGTGGAATCATGA
TACCACTTTTAGCTCTTTGCATCTTCCTTCAGTGTATTTTTGTTTTTCAAGAGG
10 AAGTAGATTTTAACTGGACAACCTTTGAGTACTGACATCATTGATAAAATAAACT
GGCTTGTGGTTTCAA (SEQ ID NO:506). Also preferred are polypeptide fragments encoded by these polynucleotide fragments. More preferably, polypeptide fragments comprise the amino acid sequence: LDELM AHLTEMQAKVAVRAD
AGKKHL PDKQDHKASLD SMLGGLEQELQDLGIATVPKGHCASCQKPLAGKVI
15 HALGQSWHPEHFVCTHCKEEIGSSPFFERSGLXYCPNDYHQLFSPRCAYCAAP
ILDKVLTAMNQ TWHEHFFCSHCGEVFGAEGFHEKDKKPYCRKDFLAMFSPK
CGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCELHYH
HRRGTLCHGCGQPITGRCISAMGYKFHPEHFVCAFCLTQLSKGIFREQNDKTY
CQPCFNKLF (SEQ ID NO:507); KASLD SMLGGLEQELQDLGIATVPKGHC
20 ASCQKPIAGKVIHAL (SEQ ID NO:508); CPNDYHQLFSPRCAYCAAPILDKVL
TAMNQ TWHEHFFCSHCGEVFGAEG (SEQ ID NO:509); DKKPYCRKDFLAM
FSPKCGGCNRPVLENYLSAMDTVWHPECFVCGDCFTSFSTGSFFELDGRPFCE
L (SEQ ID NO:510); CGQPITGRCISAMGYKFHPEHFVCAFCLTQLSKGIFRE
QNDKTYCQ (SEQ ID NO:511). Polynucleotide fragments encoding these preferred
25 polypeptide fragments are also contemplated.

This gene is expressed primarily in brain, and to a lesser extent in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
30 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disease states and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and
35 nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Moreover, since this gene shares homology with a gene that maps to chromosome 11, (See Accession No.T87404), gene as well as its translated product may be used for linkage analysis on chromosome 11.

The tissue distribution and homology to paxillin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and or detection of disease states associated with abnormal signal transduction in brain and/or the developing embryo. This would include treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in fetal spleen, brain, and to a lesser extent in six week old embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders, neurological disorders, and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 290 as residues: Arg-28 to Gly-34.

The expression of this gene in fetal spleen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. In addition the expression of this gene in the early embryo, indicates a key role in embryo development and hence the gene or gene product could be used in the treatment and or detection of embryonic development defects. This would include

treatment or detection of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntintons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder and also in the treatment and or detection of embryonic development defects.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

The translation product of this gene shares sequence homology with the gene disrupted in the neurodegenerative disease dentatorubal-pallidoluysian atrophy. Moreover a long open reading frame exists in an alternative frame. Preferred polypeptide fragments

10

comprise the following:

MGSSQSVEIPGGGTEGYHVLRVQENSPGHRAGLEPFFDFIVSINGSRLNKDND
 TLKDLLKXNVEKPKV KMLIYSSKTLELRETSVTPSNLWGGQGLLGVSIRFCSFD
 GANENVWHVLEVESNSPAALAGLRPHSDYIIGADTMNESEDLFSLIETHEAKP
 LKLYVYNTDTDNCREVIITPNSAWGGEGSLGCGIGYGYLHRIPTRPFEEGKKIS
 15 LPGQMAGTPITPLKDGFTVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVSS
 VLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLPNLP
 PHIMPGVGLPELVNPGLPPLPSMPPRNLPGLIAPLPSEFLPSFPLVPESSSAASS
 GELLSSLPPTSNAPSDPATTTAKADAASSLTVDVTPPTAKAPTTVEDRVGDSTPV
 SEKPVSAAVDANASESP (SEQ ID NO:512); SVEIPGGGTEGYHVLRVQENSPGH
 20 RAGLEPFFDFIVSINGSRLNKDNDTLKDLLKXNVEKPKV KMLIYSSKTLELRETS
 VTPSNLWGGQGLLGVSIRFCSFDGANENVWH (SEQ ID NO:513); ESNPAAL
 LAGLRPHSDYIIGADTMNESEDLFSLIETHEAKPLKLYVYNTDTDNCREVIITP
 NSAWGGEGSLGCGIGYGYLHRIPTRPFEEGKKISLPGQMAGTPITPLKDGFTV
 QLSSVNPPSLSPPGTTGIEQSLTG LSISS (SEQ ID NO:514); RIPTRPFEEGKKI
 25 SLPGQMAGTPITPLKDGFTVQLSSVNPPSLSPPGTTGIEQSLTGLSISSTPPAVS
 SVLSTGVPTVPLLPPQVNQSLTSVPPMNPATTLPGLMPLPAGLPNLPNLPNLP
 APHIMPGVGLPELVNPGLPPLPSMPPRN (SEQ ID NO:516); PGLPPLPSMPPRN
 LPGIAPLPSEFLPSFPLVPESSSAASSGELLSSLPPTSNAPSDPATTTAKADAA
 SSLTVDVTPPTAKAPTTVEDRVGDSTPVSEKPVSAAVDAN (SEQ ID NO:517).

30

This gene is expressed primarily in prostate cancer, and to a lesser extent in the pineal glands and in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological conditions and pulmonary disorders. Similarly,

35

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the nervous, pulmonary, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 291 as residues: Asn-9 to Leu-14.

The abundance of this gene in the pineal gland and its homology to a gene
10 disrupted in the neurodegenerative disease state Dentatorubral-pallidoluysian atrophy indicates that this gene may be useful in the treatment and/or detection of other neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. The abundance of this gene in fetal
15 lung would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung; that it may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis; and thus the gene or the gene protein encoded by the gene could be used in the detection and/or treatment of these
20 pulmonary disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

This gene is expressed primarily in the developing embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of
30 the above tissues or cells, particularly of the developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the
35 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene primarily in the embryo, indicates the gene plays a key role in embryo development and that the gene or the protein encoded by the gene could be used in the treatment and or detection of developmental defects in the embryo or in infants.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene displays homology to nestin, an intermediate filament protein, the expression of which correlates with the proliferation of Central Nervous System progenitor cells and that is useful in the identification of brain tumors. This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. AA527348).

This gene is expressed primarily in kidney and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders and neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the excretory and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 293 as residues: Thr-128 to Asn-135.

The tissue distribution and homology to nestin indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and/or treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, its abundance in kidney indicates that it is useful in the treatment and detection of acute renal failure and other disease states associated with the kidney.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 61

Gene shares homology with the latrophilin-related protein 1 precursor as well as the calcium-independent alpha-latrotoxin receptor. Preferred polypeptide fragments :

comprise the following amino acid sequence:

IYKVFRTAGLKPEVSCFENIRSCARXXXXXXXXXXXXXWIFGVLHVHVASVV
TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPCC (SEQ ID NO:518);
WIFGVLHVHVASVV TAYLFTVSNAFQGMFIFLFLCVLSRKIQEEYYRLFKNVPC

5 C (SEQ ID NO:519). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 2213659) The translation product of this gene shares sequence homology with CD 97, a seven transmembrane bound receptor.

This gene is expressed primarily in infant brain and in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
15 a number of disorders of the above tissues or cells, particularly of the neurological and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the
20 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 294 as residues: Lys-13 to Leu-21.

The tissue distribution of this gene suggest that it may be useful in the detection and/or treatment of neurodegenerative disease states and behavioral disorders such as
25 Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder, while its expression in hematopoietic cell types indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, asthma and immunodeficiency diseases.

30

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

This gene is expressed primarily in fetal liver and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 295 as residues: Ser-91 to Lys-98.

The tissue distribution of this gene fetal liver and spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma and immunodeficiency diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

Gene shares homology with human serum amyloid protein. Preferred polypeptide fragments comprise the following amino acid sequence:
ALTRIPPGDWVINVTAVSFAGKTTARFFHSSPPSLGDDQARTDPGHQRRD (SEQ ID NO:520) (See Accession No. W13671). Also preferred are polynucleotide fragments encoding these polypeptide fragments This gene maps to chromosome 9, and therefore, may be used as a marker in linkage analysis for chromosome 9 (See Accession No. AA004342).

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in fetal liver-spleen indicates that the gene could be important for the treatment or detection of immune or hematopoietic disorders including arthritis, leukemia, asthma, and immunodeficiency diseases.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 64

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. AA219669).

This gene is expressed specifically in the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disease states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of
15 the above tissues or cells, particularly of the neurological systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the
20 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's
25 Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Gene shares homology with a yeast protein. Preferred polypeptide fragments
30 comprise the following amino acid sequence: LQEVNITLPENSVWYERYKFDIP VFHL (SEQ ID NO:521). Also preferred are polynucleotide fragments encoding these polypeptide fragments. (See Accession No. 1332638)

This gene is expressed primarily in fetal tissue (fetus and fetal liver).

Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver disorders and cancers (e.g. hepatoblastoma). Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 298 as residues: Asn-59 to Glu-64.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

Gene has homology with a B-cell surface antigen which may indicate gene plays a role in the immune response, including, but not limited to disorders and infections of the immune system. Preferred polynucleotide fragments comprise the following sequence: TAGCATGTAGCCAGTCTGAATAACNTATAAGGACAAAGTGGAGTC CACGCGTGC GGCCGTCTAGACTAGTGGATCCCCCGGCTGCAGGATTCGGC ACGAG (SEQ ID NO:523). Also preferred are polypeptide fragments encoded by these polynucleotide fragments (See Accession No.T94535). Additionally, this gene shares homology with an interferon-gamma receptor. Preferred polypeptide fragments also comprise the following amino acid sequence: MQGSGSQFRACLLCLCFSCPC SPGGPRWNSRQGGRFPKTCRAISQNLVFKYKTFCPVRYMQPHRSSLCLHFTS YVFILSTWGLRITYSTD LKKKKKNSRGGPVPIRPKS (SEQ ID NO:522); MQGSGSQFRACLLCLCFSCPCSPGGPRWNSRQGGRFPKTCRAISQNLVFK (SEQ ID NO:524); PVRYMQPHRSSLCLHFTSYVFILSTWGLRITYSTD LKKKKK NSRGGPVPIRPKS (SEQ ID NO:525); and GEEQRDCSLGWRGVGMRATHCQAA RMFVLFSLPKYAGL (SEQ ID NO:526). Also preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in T-cells and gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders and conditions (immunodeficiencies, cancer, leukemia, hematopoiesis). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 299 as residues: Thr-41 to Gly-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders including: leukemias, lymphomas, auto-immune disorders, immunosuppressive (transplantation) and immunodeficiencies (e.g. AIDS), inflammation and hematopoietic disorders. The expression of this gene in gall bladder would suggest a possible role for this gene product in digestive disorders, particularly of the pancreas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene maps to chromosome 11, and therefore, may be used as a marker in linkage analysis for chromosome 11 (See Accession No. AA011622).

This gene is expressed primarily in a variety of fetal and developmental tissues (e.g. fetal spleen, infant brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, immune or neurological abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 300 as residues: Ser-38 to Ser-43.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for developmental abnormalities or fetal deficiencies. The detection in infant brain would suggest a role in neurological disorders (both developmental and neurodegenerative conditions of the brain and nervous system, behavioral disorders, depression, schizophrenia, Alzheimer's disease, Parkinson's
10 disease, Huntington's disease, mania, dementia). In addition, the detection in spleen would similarly suggest a role in detection and treatment of immunologically mediated disorders (e.g. immunodeficiency, inflammation, cancer, wound healing, tissue repair, hematopoiesis).

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 68

This gene is expressed primarily in spleen, T-cells, and fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
20 not limited to, immunological deficiencies, including AIDS and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and cardiovascular systems, expression of this gene at significantly higher
25 or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, autoimmune disorders, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. The expression in fetal heart indicates that polynucleotides and
35 polypeptides corresponding to this gene are useful for the treatment and diagnosis of cardiovascular disorders (e.g. heart disease, restenosis, atherosclerosis, stoke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

Gene shares homology with a human collagen protein. Preferred polypeptide fragments comprise the following amino acid sequence:

- 5 MPRKTSKCRQLLCSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPGCXSV
SSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHSKSQGE
GQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGGVKVAATTEREPEFKIK
TGKA (SEQ ID NO:527); CSGASRNADTAARQSTCSSHRPPGKIPSLGPRRXPG
CXSVSSRGEQSTGSPAAPRCGRRDAHRGLPGGAAMTPGDTWASFNPRAGHS
10 (SEQ ID NO:528); QGEGQESSGASRQDRHPVSHWVERQREAWGAPRSSSAGG
VKVAATTEREPEFKIKTGKA (SEQ ID NO:529) (See Accession No. 124886). Also
preferred are polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in fetal heart.

- Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cardiovascular disorders. Similarly, polypeptides and antibodies directed
to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
20 tissues or cells, particularly of the cardiovascular system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
25 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 302 as residues:
Pro-32 to Ser-39.

- The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for the treatment and diagnosis of cardiovascular
30 disorders (e.g. heart disease, restenosis, atherosclerosis, stroke, angina, thrombosis).

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

- The translation product of this gene shares sequence homology with a chicken
single-strand DNA-binding protein. Preferred polypeptide fragments comprise the
35 following amino acid sequence:
MSPRYPGGPRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRM
TPPRGMVPLGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNTNAN

SIPYSSASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPNR
 PNFPMPGSDGPMGGLGGMESHMHMNGSLGSGDMDSISKNSPNNMSLSNQ
 GTPRDDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:530); MSPRYPGG
 PRPPLRIPNQALGGVPGSQPLLPSGMDPTRQQGHPNMGGPMQRMTPPRGMVP
 5 LGPQNYGGAMRPPLNALGGPGMPGMNMGPGGGRPWPNTNANSIPYSSASP
 GNY (SEQ ID NO:531); LNALGGPGMPGMNMGPGGGRPWPNTNANSIPYSS
 ASPGNYVGPPGGGGPPGTPIMPSPADSTNSGDNMYTLMNAVPPGPN (SEQ ID
 NO:532); GPMGGLGGMESHMHMNGSLGSGDMDSISKNSPNNMSLSNQPGTPR
 DDGEMGGNFLNPFQSESYSPSMTMSV (SEQ ID NO:533); TCEHSSEAKAFHDY
 10 (SEQ ID NO:534). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments. (See Accession No. 1562534)

This gene is expressed primarily in placenta and to a lesser extent in the fetal heart and a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as
 15 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, developmental abnormalities, fetal deficiencies, and particularly of the
 cardiovascular system. Similarly, polypeptides and antibodies directed to these
 polypeptides are useful in providing immunological probes for differential identification
 20 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
 particularly of the reproductive system, expression of this gene at significantly higher or
 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
 another tissue or cell sample taken from an individual having such a disorder, relative to
 25 the standard gene expression level, i.e., the expression level in healthy tissue or bodily
 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for the detection and treatment of developmental
 abnormalities or fetal deficiencies, ovarian and other endometrial cancers, reproductive
 30 dysfunction, cardiovascular disorders, and pre-natal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene is expressed primarily in fetal liver and to a lesser extent in the breast and testes.

35 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, liver disorders (including hepatoblastomas) and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). The expression in testes and breast indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of endocrine and reproductive disorders (e.g. sperm maturation, milk production, testicular and breast cancers).

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. W93595).

This gene is expressed primarily in smooth muscle and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cardiovascular and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of restenosis, atherosclerosis, stroke, angina, thrombosis, wound healing and other conditions of heart disease. In addition, the expression in brain would suggest that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

Gene shares homology with human stromalin-2. Preferred polypeptide fragments comprise the following amino acid sequence:

QAFVLLS DLLLI FSPQMIVGGRDFLRPLVFFPEATLQSELASFLMDHVF IQPGDL
GSGA (SEQ ID NO:535); ACSYLLCNPEFTFFSRADFARSQLVDLLTDRFQQE
LEELLQVG (SEQ ID NO:536), QKQLSSLRDRMVAFCELCQSCLSDVDTEIQEQV
ST (SEQ ID NO:537); QVILPALTLVYFSILWTLTHISKSDAS (SEQ ID NO:538);
STHDLTRWELYECCQLLQKAVDTGXVPHQV (SEQ ID NO:539). Also preferred

are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R65208) This gene maps to chromosome 7, and therefore, may be used as a marker in linkage analysis for chromosome 7 (See Accession No. D52585).

This gene is expressed primarily in the brain (infant brain, adult brain, pituitary, cerebellum, hippocampus, schizophrenic hypothalamus, amygdala).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental and neurodegenerative diseases of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 306 as residues: Thr-25 to Lys-36, Lys-55 to Ser-63.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of developmental, degenerative and behavioral conditions of the brain and nervous system (e.g. schizophrenia, depression, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders).

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed primarily in the hypothalamus of a human suffering from schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the CNS particularly schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, such as schizophrenia expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 307 as residues: Gly-38 to Ala-44.

The tissue distribution indicates that the protein products of this gene are useful for the study, diagnosis and treatment of schizophrenia and other disorders involving the CNS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

Preferred polypeptides of the invention comprise the following amino acid sequence encoded by this gene:

LAVSTSFICCADISTALPLGSSRPAPAPRHREHEHGHQARPPRLIXTSLMPLSTP
AAQQLLWTQLTPMGGRPGGRHSPPTLHTGPRALPPGPPHPSLHVAALSLLR

(SEQ ID NO:540). Polynucleotides encoding such polypeptides are also provided.

This gene is expressed primarily in endometrial tumor and to a lesser extent in amniotic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and immune disorders particularly cancers of those systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 308 as residues: Ser-3 to Arg-9.

The tissue distribution indicates that the protein products of this gene are useful for study and treatment of immune and reproductive disorders particularly cancers of those systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in kidney cortex and to a lesser extent in early stage human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, renal disorders such as renal cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 309 as residues: Gly-38 to Gly-45, Gly-47 to Gly-52, Pro-92 to Lys-110.

The tissue distribution indicates that the protein products of this gene are useful for study, treatment and diagnosis of renal diseases such as cancer of the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in kidney medulla.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, metabolic and renal disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
10 the above tissues or cells, particularly of the metabolic and renal systems, expression of
this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
15 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that the protein products of this gene are useful
for study, treatment and diagnosis of metabolic and renal diseases and disorders.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed in chronic synovitis and microvascular endothelium.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, arthritis and atherosclerosis. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the vascular and skeletal systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

35 The tissue distribution indicates that the protein products of this gene are useful
for study, diagnosis and treatment of arthritic and other inflammatory diseases as well
as cardiovascular diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed in resting T-cells and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune system, expression of this gene at
significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for the study and treatment of immune diseases such as inflammatory conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 80

20 This gene is expressed in a variety of immune system tissues, e.g., neutrophils,
T-cells, and TNF induced epithelial and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, infectious and immune disorders. Similarly, polypeptides and antibodies
directed to these polypeptides are useful in providing immunological probes for
differential identification of the tissue(s) or cell type(s). For a number of disorders of
the above tissues or cells, particularly of the immune and vascular systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
30 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
35 NO: 313 as residues: Met-1 to Trp-6.

The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of infectious diseases, immune and vascular disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and other immune conditions. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
10 of the above tissues or cells, particularly of the immune system, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

20 This gene is expressed in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammatory and other immune conditions. Similarly, polypeptides and
25 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system, expression of this gene
at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
30 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
such a disorder, relative to the standard gene expression level, i.e., the expression level
in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
epitopes include those comprising a sequence shown in SEQ ID NO: 315 as residues:
Ala-83 to Thr-91.

35 The tissue distribution indicates that the protein products of this gene are useful
for study and treatment of immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and immune disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
10 of the above tissues or cells, particularly of the immune and inflammatory system,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
15 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

This gene is expressed in human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, disorders of the inflammatory and immune systems. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
a number of disorders of the above tissues or cells, particularly of the inflammatory and
immune systems, expression of this gene at significantly higher or lower levels may be
30 routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
cell sample taken from an individual having such a disorder, relative to the standard
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

35 The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the immune and inflammatory systems.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, inflammation and immune system diseases. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
10 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the immune system and inflammatory
system, expression of this gene at significantly higher or lower levels may be routinely
detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene
expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of diseases of the inflammatory and immune systems.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene is expressed in activated neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
25 not limited to, inflammation and immune system disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the inflammatory and immune system,
expression of this gene at significantly higher or lower levels may be routinely detected
30 in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
35 NO: 319 as residues: Met-1 to Gly-6, Gly-32 to Pro-43, Leu-55 to Gln-60.

The tissue distribution indicates that the protein products of this gene are useful
for diagnosis and treatment of disorders of the immune and inflammatory system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 87

In specific embodiments, polypeptides of the invention comprise the sequence:

EQVLALLWPRFELILEMNVQSVRSTDPQRLGGLDTRPHYITRRYAEFSSALVSIN
 5 QTIPNERTMQLLGQLQVEVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVLME
 RAADDSEVESFQQLLNARTQEFIEELLSPFGLVAFVKEAEALIERGQAERLR
 GEEARVTQLIRGFGSSWKSSVESLSQDVMRSFTNFRNGTSIIQG (SEQ ID
 NO:541), ALLKYRFFYQFLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRLMK
 VQYEEVAEKDDLGMGVEDTAKKGFXSKPSRSRNTIFTLGTRGSVISPTLEAPILV
 10 PHTAQR (SEQ ID NO: 542), EQRYPFALFRSQHYXLLDNSCREYLFICEFFVVS
 GPXAHDLFHAVMGRTLSMTLKHLD SYLADCYDAIAVFLCIHIVLRFRNIAAKRD
 VPALDRYW (SEQ ID NO:543), GGLDTRPHYITRRYAEFSSALVSINQ (SEQ ID
 NO:544), SRKEQLVFLINNYDMMLGVL (SEQ ID NO: 545) and/or ALLKYRFFY
 QFLGNERATAKEIRDEYVETLSKIYLSYYRSYLGRLMKVQYEEVAEKDDLGMG
 15 VEDTAKKGFXSKPSLSRNTIFTLGTRGSVISPTLEAPILVPHTAQRXEQRYPF
 EALFRSQHYXLLDNSCREYLFICEFFVVS GPXAHDLFHAVMGRTLSMTLKHLD
 SYLADCYDAIAVFLCIHIVLRFRNIAAKRDVPALDRYWEQVLALLWPRFELILEM
 NVQSVRSTDPQRLGGLDTRPHYITRRYAEFSSALVSINQTIPNERTMQLLGQLQV
 EVENFVLRVAAEFSSRKEQLVFLINNYDMMLGVL MERAADDSEVESFQQLLN
 20 ARTQEFIEELLSPFGLVAFVKEAEALIERGQAERLRGEEARVTQLIRGFGSSW
 KSSVESLSQDVMRSFTNFRNGTS (SEQ ID NO:546). Polynucleotides encoding

these polypeptides are also encompassed by the invention. The translation product of this gene shares sequence homology with suppressor of actin mutation which is thought to be important in mutation suppression.

25 This gene is expressed primarily in fetal liver and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are
 30 not limited to, liver and mutations. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver or cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 35 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 320 as residues: Val-53 to Arg-60, Thr-88 to Thr-94, Ala-142 to Ser-150, Gly-188 to Glu-196, Gly-208 to Ser-214, Thr-227 to Gly-232, Lys-279 to Phe-285.

- 5 The tissue distribution and homology to suppressor of actin mutation suggest that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and of liver disorder or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 88

- 10 This gene maps to chromosome 9, and therefore can be used in linkage analysis as a marker for chromosome 9. In specific embodiments, polypeptides of the invention comprise the sequence:

YEGKEFDYVFSIDVNEGGPSYKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVA
KFIIDNTKGQMLGLGNPSFSDPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYV
15 PGASMGTTMAGVDPFTGNSAYRSAASKTMNIYFPKKEAVTFDQANPTQILGK
LKELNGTAPEEKKLTEDDLILLEKILSLICNSSSEKPTVQQQLQILWKAINCPEDIV
FPALDILRLSIKHPSVNENFCNEKEGAQFSSHLINLLNPKGKPANQLLALRTFC
NCFVGQAGQKLMMSQRESLMSHAIELKSGSNKNI (SEQ ID NO: 547);
HIALATLALNYSVCFHKD (SEQ ID NO: 548); HNIEGKAQCLSLISTILEVVQ
20 DLEATFRLLVALGTLISDDSNVQLAKS (SEQ ID NO: 549); LGVDSQIKKYSS
VSEPAKVSECCRFILNLL (SEQ ID NO: 550); and/or YEGKEFDYVFSIDVNEGGPS
YKLPYNTSDDPWLTAYNFLQKNDLNPMFLDQVAKFIIDNTKGQMLGLGNPSFS
DPFTGGGRYVPGSSGSSNTLPTADPFTGAGRYVPGASMGTTMAGVDPFTGN
SAYRSAASKTMNIYFPKKEAVTFDQANPTQILGKLKELNGTAPEEKKLTEDDLI
25 LLEKILSLICNSSSEKPTVQQQLQILWKAINCPEDIVFPALDILRLSIKHPSVNENFC
NEKEGAQFSSHLINLLNPKGKPANQLLALRTFCNCFVGQAGQKLMMSQRESL
MSHAIELKSGSNKNIHIALATLALNYSVCFHKDHNIEGKAQCLSLISTILEVVQD
LEATFRLLVALGTLISDDSNVQLAKSLGVDSQIKKYSSVSEPAKVSECCRFILN
LL (SEQ ID NO: 551). Polynucleotides encoding these polypeptides are also
30 encompassed by the invention. These polypeptides share significant homology with
phospholipase A2 activating protein which is thought to be important in signal
transduction (see, e.g., Wang et al., Gene 161(2):237-241 (1995)).

- This gene is expressed primarily in endothelial cells, to a less extent in placenta, endometrial stromal cells, osteosarcoma, testis tumor, muscle, and infant brain that are
35 likely to be rich in blood vessels.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in vascular system, aberrant angiogenesis, tumor angiogenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system or tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in endothelial cells and several potential highly vascularized tissues and its homology to phospholipase A2 activating protein suggest that this gene may be involved in transducing signals for endothelial cells in angiogenesis or vasculogenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 89

In specific embodiments, polypeptides of the invention comprise the sequence: YPNQDGDILRDQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTIS AYKTPRDKVQCILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLL STVQYISSFYASCLSGEESYWWMQFTA AVE (SEQ ID NO:552); YPNQDGDILR DQVLHEHIQRLSKVVTANHRALQIPEVYLREAPWPSAQSEIRTISAYKTPRDKVQ CILRMCSTIMNLLSLANEDSVPGADDFVPVLVFLIKANPPCLLSTVQYISSFYA SCLSGEESYWWMQFTA AVEFIKTI (SEQ ID NO:553); YPNQDGDILRDQVL (SEQ ID NO:554); EAPWPSAQSEI (SEQ ID NO:555); PVLVFLIKANP (SEQ ID NO:560); SGEESYWWMQFTA AVEFIKTI (SEQ ID NO:556); ADDFVPVLVFLIKANPP (SEQ ID NO:557); YKTPRDKVQCIL (SEQ ID NO:558); and/or GADDFVPVLVFLIK (SEQ ID NO:559). The translation product of this gene shares sequence homology with human ras inhibitor and yeast VPS9p which is thought to be important in golgi vacuole transport.

This gene is expressed primarily in T cells and melanocytes and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dysfunction and disorders involving T cells and melanocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ras inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating signal transduction; diagnosis and treatment of disorders involving T cells and melanocytes.

FEATURES OF PROTEIN ENCODED BY GENE NO: 90

This gene maps to chromosome 9 and therefore polypeptides of the invention can be used in linkage analysis as a marker for chromosome 9. The translation product of this gene shares sequence homology with neuronal olfactomedin-related ER localized protein which is thought to be important in influence the maintenance, growth, or differentiation of chemosensory cilia on the apical dendrites of olfactory neurons. In specific embodiments, polypeptides of the invention comprise the sequence: SARASTQPPAGQHGPC (SEQ ID NO:561); MPGRWRWQRDMHPARKLLSLL FLILMGTELTQD (SEQ ID NO:562); SAAPDSLLRSSKGSTRGSL (SEQ ID NO:563); AAIVWIRGKSESRIAKTPGI (SEQ ID NO:564); FRGGGTLLVLPPTHT PEWLIL (SEQ ID NO:567); PLGITLPLGAPETGGGD (SEQ ID NO:565); and/or CAAETWKGSQRAGQLCALLA (SEQ ID NO:566).

This gene is expressed in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological and endocrinological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurological or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 323 as residues: Leu-20 to Ala-26, Arg-32 to Arg-39, Thr-104 to Gly-112.

The tissue distribution and homology to olfactomedin-related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for maintenance, growth, or differentiation of neuron cells in pineal gland, therefore, may be useful for diagnosis and treatment of neurological disorders in pineal gland.

FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in prostate and apoptotic T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate disease and T cell dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detect abnormal activity in prostate and T cells or probably treatment of this abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 92

This gene is expressed primarily in prostate and to a lesser extent in smooth muscle cells, fibroblasts, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in prostate or vascular system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate or vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating function of prostate or highly vascularized tissues, e.g. placenta.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

This gene is expressed primarily in embryos and fetal tissues stage human and to a lesser extent in a wide variety of other proliferative tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders in embryonic development and cell proliferation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic tissues and proliferative cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of abnormalities in developing and proliferative cells and organs.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

The translation product of this gene shares sequence homology with transformation related protein which is thought to be important in transformation.

This gene is expressed primarily in female reproductive tissues, i.e., breast cancer cells, placenta, and ovary and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancer or dysfunction of reproductive tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproduction system,
5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
10 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 327 as residues: Ser-50 to Pro-61.

The tissue distribution and homology to transformation related protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of conditions caused by transformation, i.e. tumorigenesis in
15 reproductive organs, e.g. breast, placenta, and ovary.

FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in testes, rhabdomyosarcoma, infant brain and to a lesser extent in some tumors and highly vascularized tissues.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumorigenesis, abnormal angiogenesis, and/or neurological disorders. . Similarly, polypeptides and antibodies directed to these polypeptides are useful in
25 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumor tissues or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 328 as residues: Arg-46 to Trp-54, Pro-60 to Ile-69, Asn-116 to Ala-122, Arg-147 to Lys-153, Ser-158 to Glu-170, Ile-399 to
35 Ser-405, Pro-486 to Met-499, Pro-502 to Asp-508.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for a range of disease states including treatment of

tumor or vascular disorders and the treatment of neurological disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

This gene maps to chromosome 7 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 7. The translation product of this gene is homologous to the *Clostridium perfringens* enterotoxin (CPE) receptor gene product and shares sequence homology with a human
10 ORF specific to prostate and a glycoprotein specific to oligodendrocytes both of which are tissue specific proteins. (See e.g., Katahira et al., *J Cell Biol.* 136(6):1239-1247 (1997). PMID: 9087440; UI: 97242441.

This gene is expressed primarily in pancreas tumor and ulcerative colitis and to a lesser extent in several tumors and normal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic disorder, ulcerative colitis, tumors and food poisoning. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
20 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or tumorigenic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
25 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 329 as residues: Gly-147 to Met-152, Cys-177 to Lys-188.

30 The tissue distribution and homology to prostate and oligodendrocyte-specific protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis or treatment of disorder in pancreas, ulcerative colitis, and tumors. Furthermore, identity to the human receptor for *Clostridium perfringens* enterotoxin indicates that the soluble portion of this receptor could be used in the
35 treatment of food poisoning associated with *Clostridia perfringens* by blocking the activity of *perfringens* enterotoxin.

FEATURES OF PROTEIN ENCODED BY GENE NO: 97

The translation product of this gene shares sequence homology with ATPase which is thought to be important in metabolism.

5 This gene is expressed primarily in testes and several hematopoietic cells and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 330 as residues: Leu-37 to Ala-42.

The tissue distribution and homology to ATPase indicates that polynucleotides and polypeptides corresponding to this gene are useful for marker of diagnosis and treatment of leukemia and other hematopoietic disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

In specific embodiments, polypeptides of the invention comprise the sequence: MRSARPSLGCLPSWAFSQALNI (SEQ ID NO:568); LLGLKGLAPAEISAVCE KGNFN (SEQ ID NO:569); VAHGLAWSYYIGYLRLLPELQARIR (SEQ ID NO:570); TYNQHYNLLRGAVSQRC (SEQ ID NO:571); ILLPLDCGVDPNLSM ADPNIRFLDKLPQQTGDRAGIKDRVYSN (SEQ ID NO:572); SIYELLENGQRAGT CVLEYATPLQTLFAMSQYSQAGFSGEDRLEQ (SEQ ID NO:573); AKLFCRTLE DILADAPESQNNCRLLAYQEPADDSSFSLSQEVLRHLRQEEKEEVTVGSLKTSAV PSTSTMSQEPPELLISGMEKPLPLRTDFS (SEQ ID NO:574); and/or LLGLKGLA PAEISAVCEKGNFNVAHGLAWSYYIGYLRLLPEL (SEQ ID NO:575).

35 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in prostate BPH and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, benign prostatic hypertrophy or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urinary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 331 as residues: Ile-60 to Asn-69, Leu-106 to Asp-112, Glu-130 to Gly-136, Phe-160 to Glu-167, Pro-184 to Cys-190, Glu-197 to Ser-202, Arg-215 to Glu-221, Thr-237 to Pro-242.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of benign prostatic hypertrophy or prostate cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders or injuries of the salivary gland. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders of, or injuries to the salivary gland or other glandular tissue.

FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene maps to chromosome 15, accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 15. The translation product of this gene shares sequence homology with a *C.elegans* gene of unknown function. In specific embodiments, polypeptides of the invention comprise the sequence: DPRVRLNSLTCKHIFISLTQ (SEQ ID NO:583); TMKLLKLRRNIV KLSLYRHFTN (SEQ ID NO:576); TLILAVAASIVFIWTTMKFRI (SEQ ID NO:577); VTCQSDWRELWVDDAIWRLLFISMILFVI (SEQ ID NO:578); MVLWR PSANNQRFAFSPLSEEEEEDEQ (SEQ ID NO:580); KEPMLKESFEGMKMRS TKQEPNGNSKVNKAQEDDL (SEQ ID NO:584); and/or KWVEENVPSSVTDVALP ALLDSDEERMITHFERSKME (SEQ ID NO:582). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thyroid and to a lesser extent in osteoclastoma, kidney medulla, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, thyroid dysfunction or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 333 as residues: Lys-107 to Leu-124, Glu-150 to Thr-159, Pro-173 to Asp-179, Ser-192 to Ser-201.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of thyroid dysfunction or cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 101

This gene maps to chromosome 16, therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 16. In specific embodiments, polypeptides of the invention comprise the sequence:

IRHELTVLRDTRPACA (SEQ ID NO:585); and/or MDFXMALIYD (SEQ ID NO:586). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in kidney cortex and to a lesser extent in adult brain, corpus colosum, hippocampus, and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 102

In specific embodiments, polypeptides of the invention comprise the sequence: MQEMMRNQDRALSNLESIPGGYNA (SEQ ID NO:587); LRRMYTDIQEPMLSA
 AQEQF GGNPF (SEQ ID NO:588); ASLVSNNTSSGEGSQPSRTENRDPLPNPWAP
 QT (SEQ ID NO:589); SQSSASSGTASTVGGTTGSTASGTSGQSTTAPNLVPGV
 GASMFNTPG MQSLLQQITENPQLMQNMLSAPY (SEQ ID NO:590);
 MRSMMQSLSQNPDLAAQMMLNPNLFAGNPQLQEQRQQLPTFLQQ (SEQ ID
 NO:591); MQNPDTLSAMSNPRAMQALLQIQGLQTLATEAPGLIPGFTPGLG
 ALGSTGGSSGTNGSNATPSENTSPTAGT (SEQ ID NO:592); TEPGHQQFI
 QQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALATGGDINAA
 IERLLGSQPS (SEQ ID NO:593); RNPAMMQEMMRNQDRALSNLESIPGGY
 NALRRMYTDIQEPMLSAA (SEQ ID NO:594); GNPFASLVSNNTSS (SEQ ID
 NO:595); ENRDPLPNPWA (SEQ ID NO:595); GKILKDQDTLSQHGHD (SEQ ID
 NO:597); GLTVHLVIKTQNRP (SEQ ID NO:598); SELQSQMQRQLLSNPMM
 (SEQ ID NO:599); PEISHMLNNPDIMR (SEQ ID NO:600); and/or
 RQLIMANPQMQQLIQRNP (SEQ ID NO:601). Polynucleotides encoding these

polypeptides are also encompassed by the invention.

This gene is expressed primarily in breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumor systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some types of breast cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 103

The translation product of this gene shares sequence homology with secreted serine proteases and lysozyme C precursor, which is thought to be important in bacteriolytic function. In specific embodiments, polypeptides of the invention comprise the sequence: NLCHVDCQDLLNP NLLAGIHCAKRIVS (SEQ ID NO:602); LDGFEGYSLSDWLCLAFVESKFN (SEQ ID NO:603); NENADGSFDYGLFQINSHYWCN (SEQ ID NO:604); and/or NLCHVDCQDLLNP NLLAGIHCAKRIVS (SEQ ID NO:605). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infection. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Ile-62 to Phe-70, Asn-78 to Asn-84.

The tissue distribution and homology to lysozyme C precursor indicates that polynucleotides and polypeptides corresponding to this gene are useful for boosting the monocyte-macrophage system and enhance the activity of immunoagents.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 104

This gene is expressed primarily in apoptotic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of some immune disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 105

The translation product of this gene shares sequence homology with ARI protein of Drosophila (accession 2058299; EMBL: locus DMARIADNE, accession X98309), which is thought to be important in axonal path-finding in the central nervous system. In specific embodiments, polypeptides of the invention comprise the sequence IREVNEVIQNPAT (SEQ ID NO:606); ITRILLSHFNWDKEKLMERYF DGNLEKLFA (SEQ ID NO:607); NTRSSAQDMPCQICYLNYPNSYF (SEQ ID NO:608); TGLECGHKFCMQCWSEYLTTKIMEEGMGQTISCPAHG (SEQ ID NO:614); CDILVDDNTVMRLITDSKVKLKYQHLITNSFVECNRLWKCPAPD CHHVVKVQYPDAKPV (SEQ ID NO:609); CDILVDDNTVMRLITDSK

VKLKYQHLITNSFVECNRLKWCAPDCHHVVKV (SEQ ID NO:610);
 GCNHMVCRNQNCCKAEFCWVCLGPWEPHGSAWYNCNRYNEDDAKAARDAQE
 RSRAALQRYL (SEQ ID NO:611); FYCNR YMNHMQSLRFEHKLYAQVKQ
 KMEEMQQHNMSWIEVQFLKKA VDVLCQCRATLMT (SEQ ID NO: 612);
 5 YVFAFYLLKKNQSIIFENNQADLENATEVLSGYLERDISQDSLQDIKQKVQDKY
 RYCESR (SEQ ID NO:613) Polynucleotides encoding these polypeptides are also
 encompassed by the invention.

This gene is expressed primarily in adult brain, and to a lesser extent in
 endometrial tumor, melanocytes, and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases or injuries involving axonal path development. Similarly,
 polypeptides and antibodies directed to these polypeptides are useful in providing
 15 immunological probes for differential identification of the tissue(s) or cell type(s). For
 a number of disorders of the above tissues or cells, particularly of the central nervous
 system, expression of this gene at significantly higher or lower levels may be routinely
 detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g.,
 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
 20 taken from an individual having such a disorder, relative to the standard gene
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an
 individual not having the disorder.

The tissue distribution and homology to ARI protein indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for treatment of
 25 disease states or injuries involving axonal path development, including
 neurodegenerative diseases and nerve injury.

FEATURES OF PROTEIN ENCODED BY GENE NO: 106

The translation product of this gene shares sequence homology with cytochrome
 30 b561 [Sus scrofa] which is thought to be an integral membrane protein of
 neuroendocrine storage vesicles of neurotransmitters and peptide hormones.

This gene is expressed primarily in frontal cortex and to a lesser extent in
 rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
 35 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 339 as residues: Ser-18 to Pro-24.

The tissue distribution and homology to cytochrome b561 [*Sus scrofa*] indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of neurological disorders. This gene may also be important in regulation of some types of cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 107

In specific embodiments, polypeptides of the invention comprise the sequence: MWGYLFVDAAWNFLGCLICGW (SEQ ID NO:615); MHFISSGNVSAIRSSILLRXSLSYLGNCRLRVSAIFVYFLLFLLLS (SEQ ID NO:616); and/or MDQALRGSPSEGFSTDPSPPPQVGRQIPSFPPWRRLVLPKASGCFLEREWLVCVKLRTRPGAEAHAYNSSILGGRGKGIT (SEQ ID NO:617). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pancreas tumor and to a lesser extent in cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pancreatic tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 340 as residues: Pro-22 to Phe-33.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pancreatic tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene maps to chromosome 17 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

10 MLPALASCCHFSPPEQAARLKKLQEQEKQQKVEFRKRMEKEVSDFIQDSGQIK
KKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYVMIFKKEFAPSDEELDSY
RRGEEWDPQKAEEKRNXXKELAQRR (SEQ ID NO:618); EEEAAQQGPVVV
SPADYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE
IRAKKRLRQSGE (SEQ ID NO:619); PPRPAQLPLTPGAGQGAGRDKAAAIRA
15 HPGAPPLNHLPL (SEQ ID NO:620); AVPQAGGKQVFDLSPLELG YVRGMCVCV
(SEQ ID NO:621) and/or MLPALASCCHFSPPEQAARLKKLQEQEKQQKVEFRK
RMEKEVSDFIQDSGQIKKKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYV
MIFKKEFAPSDEELDSYRRGEEWDPQKAEEKRNXXKELAQRR EEEAAQQGPVVV
SPADYKDKYSHLIGKGAAKDAAHMLQANKTYGCXPVANKRDTRSIEEAMNE
20 IRAKKRLRQSGE (SEQ ID NO:622). Polynucleotides encoding these polypeptides
are also encompassed by the invention. The translation product of this gene shares
sequence homology with FSA-1 which may play a role as a structural protein
component of the acrosome.

This gene is expressed primarily in fetal kidney and sperm.

25 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, male reproductive disorders, especially involving acrosomal dysfunction.
Similarly, polypeptides and antibodies directed to these polypeptides are useful in
30 providing immunological probes for differential identification of the tissue(s) or cell
type(s). For a number of disorders of the above tissues or cells, particularly of the male
reproductive system, expression of this gene at significantly higher or lower levels may
be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
35 cell sample taken from an individual having such a disorder, relative to the standard
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an

individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 341 as residues: Glu-8 to Asn-35.

The tissue distribution and homology to FSA-1 indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of infertility due to acrosomal disfunction of sperm.

FEATURES OF PROTEIN ENCODED BY GENE NO: 109

This gene is expressed primarily in pituitary and to a lesser extent in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, male reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 342 as residues: Met-1 to Trp-6.

Because the gene is found in both pituitary and epididymus, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of male reproductive disorders. This may involve a secreted peptide produced in the pituitary targeting the epididymus.

FEATURES OF PROTEIN ENCODED BY GENE NO: 110

In specific embodiments, polypeptides of the invention comprise the sequence: LLCPVLSNGXSWNFPHP SQPEYSFHGFHSTRLWI (SEQ ID NO:623); and/or PSTPWFLFLLGLTCPFSTSHPRWDSIPP (SEQ ID NO:624). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in resting T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, T-cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of certain immune disorders, especially those involving T-cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 111

This gene is expressed primarily in cerebellum and whole brain and to a lesser extent in infant brain and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 344 as residues: Asp-48 to Gly-55.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 112

The translation product of this gene shares sequence homology with yeast mitochondrial ribosomal protein homologous to ribosomal protein s15 of E.coli which

is thought to be important in the early assembly of ribosomes (See Accession No. M38016). This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in developmental tissues.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development of cancers and tumors in addition to healing wounds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in
10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
15 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribosomal protein s15 of *E. coli* indicates that polynucleotides and polypeptides corresponding to this gene are useful for
20 diseases related to the assembly of ribosomes in the mitochondria which is important in the translation of RNA into protein. Therefore, this indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of multiple tumors as well as in healing wounds which are thought to be under similar regulation as developmental tissues. Protein, as well as, antibodies directed against the
25 protein have utility as tumor markers, in addition to immunotherapy targets, for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 113

The translation product of this gene shares sequence homology with human
30 poliovirus receptor precursors which are thought to be important in viral binding and uptake. Preferred polypeptide fragments comprise the following amino acid sequence:
ELISISINVALADEGEYTCISFTMPVRTAKSLVTVLGIPQKPIITGYKSSLREKDT
ATLNCQSSGSKPAARLTWRKGDQELHGEPTRIQEDPNGKTFTVSSSVTFQVTR
EDDGASIVCSVNHESLKGADRSTSQRIEVLYTPTAMIRPDPPHPREGQKLLHHC
35 EGRGNPVPQQYLWEKEGSPPLKMTQESALIFPFLNKSDSGTYGCTATSNMGS
YKAYYTLNVND (SEQ ID NO:625). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gnllPID1002627).

This gene is expressed almost exclusively in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, susceptibility to viral disease and diseases of the CNS especially cancers of that system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 346 as residues: Leu-26 to Asp-37, Lys-53 to Ser-59.

The tissue distribution and homology to poliovirus receptor precursors indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and prevention of diseases that involve the binding and uptake of virus particles for infection. It might also be helpful in genetic therapy where the goal is to insert foreign DNA into infected cells. With the help of this protein, the binding and uptake of this foreign DNA might be aided. In addition, it is expected that over expression of this gene will indicate abnormalities involving the CNS, particularly cancers of that system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene shares sequence homology with YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* in addition to alpha-1 collagen type III (See Accession No. gil537432). One embodiment for this gene is the polypeptide fragment(s) comprising the following amino acid sequence: VPELPDRVHQLHQA VQG CALGRPGFPGGPTH SGHHKSHPGPAGGDYNRCDRPGQVHLHNPRGTGRRGQLHPTAGPGVHRRACPSQQLPHRLGPGVPCSPSLTPVLPSWTQSWCG LPGYTSSS (SEQ ID NO:630). An additional embodiment is the polynucleotide fragment(s) encoding these polypeptide fragments

This gene is expressed primarily in brain cells and to a lesser extent in activated B and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegeneration and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 347 as residues: Glu-34 to Glu-39, Gly-51 to Ser-72, Ala-88 to Glu-93, Gln-100 to Val-105.

The tissue distribution and homology to YO87_CAEEL hypothetical 28.5 KD protein ZK1236.7 in chromosome III of *Caenorhabditis elegans* as well as to a conserved alpha-1 collagen type III protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons' Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 115

The translation product of this gene shares sequence homology with alpha 3 type IX collagen which is thought to be important in hyaline cartilage formation via its ability to uptake inorganic sulfate by cells (See Accession No. gil975657). One embodiment of this gene is the polypeptide fragment comprising the following amino acid sequence: SLRRPRSAAXQTLTTLSSVSSASSALPGSREPCDPRAPPPR SGSAASCCSCCCSCPRRRAPLRSPRGSKRRIRQREVVDLYNGMCLQGPAGVPG RDGSPGANGIPGTPGIPGRDGFKGEGECLRESFEESWTPNYKQCSWSSLNY GIDLGKIAECTFTKMRSNSALRVLFSGSLRLKCRNACCQRWYFTFNGAECSPG LPIEAIYLDQGSPEMNSTINIHRTSSVEGLCEGIGAGLVDVAIWVGTCSDYPKG DASTGWNSVSRIIIIEELPK (SEQ ID NO:634). An additional embodiment are the

polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in smooth muscle and to a lesser extent in synovial tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e., spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to alpha 3 type IX collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases associated with the mutation in this gene which leads to the many different types of chondrodysplasias. By the use of this product, the abnormal growth and development of bones of the limbs and spine could be routinely detected or treated in utero since the protein or muteins thereof could affect epithelial cells early in development and later the chondrocytes of the developing craniofacial structure.

FEATURES OF PROTEIN ENCODED BY GENE NO: 116

The translation product of this gene shares sequence homology with retrovirus-related reverse transcriptase which is thought to be important in viral replication. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: TKKENC RPASLMNIDTKILNKILMNQ (SEQ ID NO:640). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. pirlA25313IGNHUL1).

This gene is expressed primarily in human meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, retroviral diseases such as AIDS, and possibly certain cancers due to transactivation of latent cell division genes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to retrovirus-related reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of diseases and maladies associated with retroviral infection since a functional reverse transcriptase (RT) or RT-like molecule is an integral component of the retroviral life cycle.

FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with an unknown gene from *C. elegans*, as well as weak homolog with mammalian metaxin, a gene contiguous to both thrombospondin 3 and glucocerebrosidase, is known to be required for embryonic development. Preferred polypeptide fragments comprise the following amino acid sequence: MCNLPIKVVCRAEYMSPSGKVPXXHVGNGQ VVSELGPIVQFVKAKGHSLSGLEEVQKAEMKAYMELVNNMLLTAELYLQWC DEATVGXITHXRYGSPYPWPLXHILAYQKQWEVKRKXKAIGWGKKTLTDQVLE DVDQCCQALSQRLGTQPYFFNKQPTELDALVFGHL YTLTTQLTNDELSEKVKN YSNLLAFCRRI EQHYFEDRGKGRLS (SEQ ID NO:641); MCNLPIKVVCRAEYMSPSGKVPXXHVGNGQVVSELGPIVQFVK (SEQ ID NO:642). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. gil1326108).

This gene is expressed primarily in fetal tissues and to a lesser extent in hematopoietic cells and tissues, including spleen, monocytes, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer; lymphoproliferative disorders; inflammation; chondrosarcoma, and Gaucher disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification

of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and embryonic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 118

The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA chain from an RNA molecule, and is a method whereby the infecting RNA chains of retroviruses are transcribed into their DNA complements. One embodiment for this gene is the polypeptide fragment comprising the following amino acid sequence:
 MXXXNSHITIFTLVNGLNAPNERHRLANWIQSQDVCCIQETHLTGRDTHRL
 KIKGWRKIYQANGKQKK (SEQ ID NO:647). An additional embodiment is the polynucleotide fragments comprising polynucleotides encoding these polypeptide fragments (See Accession No. gil2072964).

This gene is expressed primarily in skin and to a lesser extent in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, hematopoietic disorders; inflammation; disorders of immune surveillance. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the epidermis and/or hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and

wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to reverse transcriptase indicates that polynucleotides and polypeptides corresponding to this gene are useful for cancer therapy. Expression in the skin also indicates that this gene is useful in wound healing and fibrosis. Expression by neutrophils also indicates that this gene product plays a role in inflammation and the control of immune surveillance (i.e. recognition of viral
10 pathogens). Reverse transcriptase family members are also useful in the detection and treatment of AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 119

15 The translation product of this gene shares sequence homology with reverse transcriptase which is important in the synthesis of a cDNA copy of an RNA molecule, and is a method whereby a retrovirus reverse-transcribes its genome into an inheritable DNA copy.

 This gene is expressed primarily in the frontal cortex of brain.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
25 of the above tissues or cells, particularly of the CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
30 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

 The tissue distribution and homology to reverse transcriptase suggest that this is useful in the treatment of cancer and AIDS. The expression in brain indicates that it plays a role in neurodegenerative disorders and in neural degeneration.

35

FEATURES OF PROTEIN ENCODED BY GENE NO: 120

One embodiment of this gene has homology to a hypothetical protein in *Schizosaccharomyces pombe* (See Accession No. 2281980). Another embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

5 IYHLHSWIFFHFKRAFCMCFITMKVIHAHCSKLRKCXNAQISVFCTTLTASYPT
(SEQ ID NO:651). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

10 This gene is expressed primarily in adult hypothalamus and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative disorders; endocrine function; and vertigo. Similarly,
15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS and peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
25 corresponding to this gene are useful for the treatment and diagnosis of neurodegenerative disorders; diagnosis of tumors of a brain or neuronal origin; treatments involving hormonal control of the entire body and of homeostasis, behavioral disorders, such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and
30 panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 121

35 The translation product of this gene shares sequence homology with the human IRLB protein which is thought to be important in binding to a c-myc promoter element and thus regulating its transcription (See Accession No. gi33969). This gene maps to

chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in brain and breast and to a lesser extent in a variety of hematopoietic tissues and cells.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer of the brain and breast; lymphoproliferative disorders; neurodegenerative diseases. Similarly, polypeptides and antibodies directed to these
10 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS, breast, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
15 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancer of the
20 brain, breast, and hematopoietic system. In addition, it may be useful for the treatment of neurodegenerative disorders, as well as disorders of the hematopoietic system, including defects in immune competency and inflammation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 122

The translation product of this gene shares sequence homology with an ATP synthase, a key component of the proton channel that is thought to be important in the translocation of protons across the membrane.

30 This gene is expressed primarily in T cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, T cell lymphoma. Similarly, polypeptides and antibodies directed to these
35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATP synthase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of defects in proton transport, homeostasis, and metabolism, as well as the diagnosis and treatment of lymphoma. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia

FEATURES OF PROTEIN ENCODED BY GENE NO: 123

This gene maps to chromosome 15, and therefore, may be used as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in a variety of fetal tissues, including fetal liver, lung, and spleen, and to a lesser extent in a variety of blood cells, including eosinophils and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer (abnormal cell proliferation); T cell lymphomas; and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetus and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions involving cell proliferation. Expression of this gene in fetal tissues, as well as in a variety of blood cell lineages indicates that it may play a role in either cellular proliferation; apoptosis; or cell survival. Thus it may be useful in the management and

treatment of a variety of cancers and malignancies. In addition, its expression in blood cells suggest that it may play additional roles in hematopoietic disorders and conditions, and could be useful in treating diseases involving autoimmunity, immune modulation, immune surveillance, and inflammation..

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in placenta and to a lesser extent in pineal gland and rhabdomyosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental, endocrine, and female reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For
15 a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
20 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 357 as residues: Leu-69 to Val-76.

The tissue distribution indicates that polynucleotides and polypeptides
25 corresponding to this gene are useful for diagnosis and treatment of disorders in development. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in benign prostatic hyperplasia.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of benign prostatic hyperplasia. Similarly,
35 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive

system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of benign prostatic hyperplasia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in apoptotic T-cells and to a lesser extent in suppressor T cells and ulcerative colitis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving premature apoptosis, and immunological and gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders involving inappropriate levels of apoptosis, especially in immune cell lineages. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in Raji cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and T cell autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 360 as residues: Asp-23 to Gly-29.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammation and T cell autoimmune disorders. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 128

The translation product of this gene shares sequence homology with an *C. elegans* coding region C47D12.2 of unknown function (See Accession No. gnllPIDle348986). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: EDDGFNRSIHEVILKNITWY
SERVLTEISLGSLLILVVIRTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQY
AAQRIISLFSLLSKKHNVLEQATQSLRGSLSNDVPLPDYAQDLNVIEEVIRMM
LEIINSCLTNSLHHNPNLVALLYKRDLEQFRTHPSFQDIMQNIDLVISFFSSRL
QAGS (SEQ ID NO:657); EDDGFNRSIHEVILKNITWYSERVLTEISLGSLLILVV
(SEQ ID NO:658); RTIQYNMTRTRDKYLHTNCLAALANMSAQFRSLHQYAAQ
RIISLFSLLSKKHNVLEQATQSLRGSLSNDVPLPDYAQD (SEQ ID NO:661); SCLTNSLHHNPNLVALLYKRDLEQFRTHPSFQD
IMQNIDLVISFFSSRLQAGS (SEQ ID NO:660). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to

chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

This gene is expressed primarily in smooth muscle and to a lesser extent in fetal liver.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, atherosclerosis and other cardiovascular and hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample
15 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of circulatory system
20 disorders such as atherosclerosis, hypertension, and thrombosis. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the
25 expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 129

30 The translation product of this gene shares sequence homology with a ribosomal protein which is thought to be important in cellular metabolism, in addition to the *C.elegans* protein F40F11.1 which does not have a known function at the current time (See Accession No. gnllPIDle244552). Preferred polypeptide fragments comprise the following amino acid sequence:

35 MADIQTERAYQKQPTIFQNKKRVLLGETGKEKLPRVTNKNIGLGFKDT
PRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQDEDAEDHCHPPRLSALHPQVQ
PLREAPQEHVCTPVPL LQGRPDR (SEQ ID NO:662): MKMQRTTIVIRRDYLH

YIRKYNRFEKRHKNMSVHLSPCFRDVQIGDIVTVGECRPLSKTVRFNVLKVTK
 AAGTKKQFQKF (SEQ ID NO:663); MADIQTERAYQKQPTIFQNKKRVLLGET
 GK (SEQ ID NO:664); HCHPPRLSALHPQVQPLREAPQEHVCTPVPL LQGRPDR
 (SEQ ID NO:666); NIGLGFKDTPRLLRGTYIDKKCPFTGNVSIRGRILSGVVTQ
 5 (SEQ ID NO:669); MKMQRTIVIRRDYLYIRKYNRFEKRHKNMSVHLSP (SEQ
 ID NO:667); CFRDVQIGDIVTVGECRPLSKTVRFNVLKVTKAAGTKKQFQKF
 (SEQ ID NO:668). Also preferred are polynucleotide fragments encoding these
 polypeptide fragments.

10 This gene is expressed primarily in Wilm's tumor and to a lesser extent in
 thymus and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, diseases affecting RNA translation. Similarly, polypeptides and
 15 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the Wilm's tumors, expression of this gene
 at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
 20 fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 such a disorder, relative to the standard gene expression level, i.e., the expression level
 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred
 epitopes include those comprising a sequence shown in SEQ ID NO: 362 as residues:
 Thr-11 to Asp-20.

25 The tissue distribution and homology to a ribosomal protein indicates that
 polynucleotides and polypeptides corresponding to this gene are useful for diseases
 affecting RNA translation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 130

30 The translation product of this gene shares sequence homology with a yeast
 DNA helicase which is thought to be important in global transcriptional regulation (See
 Accession No. gnllPIDle243594). One embodiment for this gene is the polypeptide
 fragments comprising the following amino acid sequence: IFYDSDWNPTVDQQA
 MDRAHRLGQTKQVTYRLLICKGTIEERILQRAKEKSEIQRMVISG (SEQ ID
 35 NO:670); TRMIDLLEEYMVYRKHTYXRLDGSSKISERRDMVADFQNRNDI
 FVFLSTRAGGLGINLTAXDTVHF (SEQ ID NO:671); TRMIDLLEEYMVYRK
 HTYXRLDGSSKISERRDM (SEQ ID NO:674); RRDMMVADFQNRNDIFVLL

STRAGGLGINLTAXDTVHF (SEQ ID NO:675) , IFYDSOWNPTVDQQAMD
RAHRLGQTKQVTYRLICKG (SEQ ID NO:676); RLICKGTIEERILQRAK
EKSEIQRMVISG (SEQ ID NO:678). An additional embodiment is the polynucleotide
fragments encoding these polypeptide fragments.

5 This gene is expressed primarily in amygdala.

 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, diseases and disorders of the brain. Similarly, polypeptides and
10 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the central nervous system, expression of
this gene at significantly higher or lower levels may be routinely detected in certain
tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
individual having such a disorder, relative to the standard gene expression level; i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

 The tissue distribution and homology to a DNA helicase indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diseases
affecting RNA transcription, particularly developmental disorders and healing wounds
since the later are though to approximate developmental transcriptional regulation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 131

25 This gene is expressed primarily in prostate and to a lesser extent in amygdala
and pancreatic tumors.

 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
30 not limited to, prostate enlargement and gastrointestinal disorders, particularly of the
pancreas and gall bladder. Similarly, polypeptides and antibodies directed to these
polypeptides are useful in providing immunological probes for differential identification
of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,
particularly of the reproductive system, expression of this gene at significantly higher or
35 lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of prostate diseases, including benign prostatic hyperplasia and prostate cancer. In addition, the tissue distribution in tumors of the pancreas indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in adult lung and to a lesser extent in hypothalamus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, pulmonary diseases and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and respiratory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of pulmonary and respiratory disorders such as emphysema, pneumonia, and pulmonary edema and emboli. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 133

5 This gene is expressed primarily in human liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cirrhosis of the liver and other hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver disorders such as cirrhosis, jaundice, and Hepatitis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 134

This gene is expressed primarily in fetal kidney and to a lesser extent in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, development and regeneration of liver and kidney and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and excretory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 367 as residues: Pro-70 to Arg-77, Tyr-102 to Thr-107.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the kidney and liver, such as cirrhosis, kidney failure, kidney stones, and liver failure, hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 135

This gene is expressed primarily in brain, bone marrow, and to a lesser extent in placenta, T cell, testis and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurodegenerative and immunological diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 368 as residues: Met-1 to His-6.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also

play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 136

Translation product of this gene is homologous to the human WD repeat protein HAN11. Preferred polypeptide fragments comprise the following amino acid sequence:

MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLALGSFVEEYNNKVQLVG
LDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDYLRVWRVGETET
RLECLLNNNKNSDFCAPLTSFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRV
NLVSGHVKTQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEH
STIIYEDPQHHPLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTIE
HVSMAILGPHIHPATSALQRM TTRLSSGTSSKCPEPLRTL SWPTQLXGEINNVQ
WASTQPELSPSATTAWRYSECSVGGAVPTRQGLLYFLPLPHPQS (SEQ ID
NO:679); MSLHGKRKEIYKYEAPWTVYAMNWSVRPDKRFRLALGSFV
EEYNNKVQLVGLDEESSEFICRNTFDHPYPTTKLMWIPDTKGVYPDLLATSGDY
LRVWRVGETETRLECLLNNNKNSDFCAPLTSFDWNEVDPYLL (SEQ ID
NO:680); SFDWNEVDPYLLGTSSIDTTCTIWGLETGQVLGRVNLVSGHVK
TQLIAHDKEVYDIAFSRAGGGRDMFASVGADGSVRMFDLRHLEHSTIIYEDPQH
HPLLRLCWNKQDPNYLATMAMDGMEVVILDVRVPAHLXPGTTI (SEQ ID
NO:681); VGADGSVRMFDLRHLEHSTIIYEDPQHHPLLRLCWNKQDPNYLA
TMAMDGMEVVILDVRVPAHLXPGTTIEHVSMAILGPHIHPATSALQRM TTRL
SGTSSKCPEPLRTL SWPTQLXGEINNVQWASTQPELSPSATTAWRYSECSV
GAVPTRQGLLYFLPLPHPQS (SEQ ID NO:682). Also preferred are polynucleotide
fragments encoding these polypeptide fragments.

This gene is expressed primarily in placenta, embryo, T cell and fetal lung and to a lesser extent in endothelial, tonsil and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological and developmental diseases in addition to cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or

cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 369 as residues: Gly-19 to Gln-28, Pro-36 to Phe-42.

5 The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above
10 listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

 This gene is expressed primarily in TNF and INF induced epithelial cells, T cells and kidney.

 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory conditions particularly inflammatory reactions in the kidney. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of renal
25 system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 370 as residues: Thr-67 to Gly-72, Gln-132 to Ala-145, Arg-150 to Pro-157.

 The tissue distribution indicates that the protein products of this gene are useful for treating the damage caused by inflammation of the kidney.

FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1 (See Accession No. D63485).

5 This gene is expressed primarily in breast cancer and colon cancer and to a lesser extent in thymus and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers, especially of the breast and colon tissues. Similarly,
10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
15 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon and breast origins indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene maps to chromosome 17, and therefore, can be used as a marker for linkage analysis from chromosome 17.

This gene is expressed primarily in CD34 positive cells, and to lesser extent in activated T-cells and neutrophils.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunologically related diseases and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoietic system, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in CD34, T-cell and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of hematopoietic disorders and immunologically related diseases, such as anemia, leukemia, inflammation, infection, allergy, immunodeficiency disorders, arthritis, asthma, immune deficiency diseases such as AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene was recently cloned by another group, who called the gene KIAA0313 gene. (See Accession No. d1021609.) Preferred polypeptide fragments comprise the amino acid sequence:

LYATATVISSPSTEXLSQDQGDRA SLDAADSGRGSWTSCSSGSHDNIQTIQ
HQRSWETLPFGHTHFDYSGDPAGLWASSSHMDQIMFSDHSTKYNRQNSRES
LEQAQSRASWASSTGYWGEDSEGDTGTIKRRGGKDVSI EAESSSLTSVTTEETK
PVPMPAHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITDFPEGHSH PARKP
PDYNVALQR SRMVARSSDTAGPSSVQQPHGHPTSSRPV NKPQWHKXNESDPR
LAPYQSQGFSTEEDEDEQVSAV (SEQ ID NO:683); HMDQIMFSDHSTKYNRQ
NQSRESLEQAQSRASWASSTGYWGE (SEQ ID NO:684); SVTTEETKPVPMP
AHIAVASSTTKGLIARKEGRYREPPPTPPGYIGIPITD (SEQ ID NO:685); and
VALQR SRMVARSSDTAGPSSVQQPHGHPTSSRPV NKPQW

HKXNESDPR LAPYQSQGF (SEQ ID NO:686). Also preferred are polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 4, and therefore, may be used as a marker in linkage analysis for chromosome 4 (See Accession No. AB002311).

This gene is expressed primarily in ovarian cancer, tumors of the Testis, brain, and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, ovarian, testicle, brain and colon cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male and female reproductive systems.

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, testis, and brain origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 141

This gene is expressed primarily in spleen and colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, colon cancer and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal tract and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of colon, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 142

Translation product is homologous to T cell translocation protein, a putative zinc finger factor (See Accession No. 340454), as well as to the G-protein coupled receptor TM5 consensus polypeptide (See Accession No. R50734). Preferred polypeptide

5 fragments comprise the following amino acid sequence:

CLLFV FVSLGMRCLFWTIVYNVLYLKHKCNTVLLCYHLCSI (SEQ ID NO:687);
ACSKLIPAFEMVMRAKDNVYHLDCFACQLCNQRXCVGDKFFLKNNXXLCQT
DYEEGLMKEGYAPXVR (SEQ ID NO:688). Also preferred are polynucleotide
fragments encoding these polypeptide fragments.

10 This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological disorders including brain cancer. Similarly, polypeptides
15 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central Nervous System, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
20 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
25 corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with
30 the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 143

Translation product for this gene has significant homology to the Fas ligand, which is a cysteine-rich type II transmembrane protein/tumor necrosis factor receptor
35 homolog. Mutations within this protein have been shown to result in generalized lymphoproliferative disease leading to the development of lymphadenopathy and autoimmune disease (See Medline Article No. 94185175). Preferred polypeptide

fragments comprise the following amino acid sequence:

SALSEPGAPDRRRPCPESVPRRPDDEQWPPPTALCLDVAPLPPSS (SEQ ID NO:689). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. 473565).

5 This gene is expressed primarily in osteoblasts, lung, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, osteoblast-related, pulmonary, neurological, and immunological
10 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
15 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 376 as residues: Trp-33 to Thr-40, Lys-
20 45 to Ile-63.

The tissue distribution in osteoblasts, lung, and brain combined with its homology to the Fas ligand indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as,
25 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the Fas ligand gene is known to be expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including asthma, immune deficiency diseases such as AIDS
30 and leukemia, and various autoimmune disorders including lupus and arthritis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene shares sequence homology with a 21.5 KD transmembrane protein in the SEC15-SAP4 intergenic region of yeast. (See Accession No. 1723971.) Preferred
35 polypeptide fragments comprise the amino acid sequence:

AHASESGERWWACCGVRFGRLRSIEAIGRSCCHDGPGLVANRGRRFKWAIEL
SGPGGSRGRSDRGSGQGDSLYPVGYLDKQVPDTSVQETDRILVEKRCWDIAL

GPLKQIPMNLFI MYMAGNTISIFPTMMVCMMAWRPIQALMAISATFKMLESSSQ
KFLQGLVYLIGNLMGLALAVYKCQSMGLLPTHASDWLAFIEPPERMEFSGG
GLLL (SEQ ID NO:691); PVGYLDKQVPDTSVQETDRILVEKRCWDIALGPLKQ
IPMNLFI (SEQ ID NO:693); and ATFKMLESSSQKFLQGLVYLIGNLMGLALAV
5 YKCQSMGLLPTHASD (SEQ ID NO:692). Also preferred are polynucleotide
fragments encoding these polypeptide fragments.

This gene is expressed primarily in osteoclastoma, hemangiopericytoma, liver,
lung.

Therefore, polynucleotides and polypeptides of the invention are useful as
10 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, osteoclastoma, hemangiopericytoma, liver and lung tumors. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the above tissue(s) or cell
15 type(s). For a number of disorders of the above tissues or cells, particularly of the lung
and liver systems, expression of this gene at significantly higher or lower levels may be
routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily
fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
cell sample taken from an individual having such a disorder, relative to the standard
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for diagnosing osteoclastoma,
hemangiopericytoma, liver and lung tumors.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 145

Translation product of this gene shares homology with the glucagon-69 gene
which may indicate this gene plays a role in regulating metabolism. (See Accession No.
A60318) One embodiment for this gene is the polypeptide fragments comprising the
30 following amino acid sequence:
PTTKLDIMEKKKHIQIRFPSFYHKLVDSEGRMRSKRETRREDSDTKHNL (SEQ ID
NO:694). An additional embodiment is the polynucleotide fragments encoding these
polypeptide fragments.

This gene is expressed primarily in brain, kidney, colon, and testis.

35 Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, brain, kidney, colon, and testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, neurological, circulatory, and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of brain, kidney, colon, and testis origins, indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 146

The translation product of this gene shares sequence homology with goliath protein which is thought to be important in the regulation of gene expression during development. Protein may serve as a transcription factor. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

TEHILAVMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGSLVFVSISFIV
LMISSAWLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTVKKGDKETD
PDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDPWLSEHCTCPMCKLNLKA
LGIV (SEQ ID NO:695); TEHILAVMITELRGKDILSYLEKNISVQMTIAVGTRMP
PKNFSRGSLVFVSISFIVLM ISSAWLIFYF (SEQ ID NO:697); SISFIVLMISSA
WLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTVKKGDKE (SEQ ID
NO:698); VKKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHVFHKSCVDP

WLSEHCTCPMCKLNILKALGIV (SEQ ID NO:699). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments (See Accession No. 157535). Moreover, another embodiment is the polynucleotide fragments encoding these polypeptide fragments:

- 5 MTHPGTEHIIA VMITELRGKDILSYLEKNISVQMTIAVGTRMPPKNFSRGS
LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR LGDAAKKAISKLTTRTV
KKGDKETDPDFDHCAVCIESYKQNDVVRILPCKHV FHKSCVDPWLSEHCTCP
MCKLNILKALGIVPNLPCTDNVAFD MERLTRTQAVNRRSALGDLAGDNSLGLE
PLRTSGISPLPQDGELTPRTGEINIAVTKEWFILASFGLLSALTLCYMIIRATASLN
10 ANEVEWF (SEQ ID NO:696); MTHPGTEHIIA VMITELRGKDILSYLEKNISVQM
TIAVGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTNARDRNQRR
LGDAAKKAISKLTTRT (SEQ ID NO:700); AAKKAISKLTTRTVKKGDKE
TDPDFDHCAVCIESYKQNDVVRILPCKHV FHKSCVDPWLSEHCTCPMCKLNIL
KALGIVPNLPC (SEQ ID NO:701); TQAVNRRSALGDLAGDNSLGLEPLRTSGI
15 SPLPQDGELTPRTGEINIAVTKEWFILASFGLLSALTLCYMIIRATASLNANEVEW
F (SEQ ID NO:702); PLHGVADHLGCDPQTRFFVPPNIKQWIAL LQRGNCTF
KEKISR AAFHNAVAVVIYNNKSKEEPVTMTHPGTEHIIA VMITELRGKDILSYLE
KNISVQMTIAVGTRMPPKNFSRGS LVFVSISFIVLMISSAWLIFYFIQKIRYTN
RDRNQRR LGDAAKKAISKLTTRTVKKGDKETDPDFDHCAVCIESYKQNDVVRIL
20 LPCKHV FHKSCVDPWLSEHCTCPMCKLNILKALGIVPNLPCTDNVAFD MERLT
RTQAVNRRSALGDLAGDNSLGLEPLRTSGISPLPQDGELTPRTGEINIAVTKEW
FILASFGLLSALTLCYMIIRATASLNANEVEWF (SEQ ID NO:703); and
HGVADHLGCDPQTRFFVPPNIKQWIAL LQRGNCTFKEKISR AAFHNAVAVVIY
NNKSKEE (SEQ ID NO:704). An additional embodiment is the polynucleotide
25 fragments encoding these polypeptide fragments. When tested against Jurkat cell lines,
supernatants removed from cells containing this gene activated the GAS pathway.
Thus, it is likely that this gene activates immune cells through the JAKS/STAT signal
transduction pathway.

- 30 This gene is expressed primarily in macrophage, breast, kidney and to a lesser
extent in synovium, hypothalamus and rhabdomyosarcoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, schizophrenia and cancer. Similarly, polypeptides and antibodies directed
35 to these polypeptides are useful in providing immunological probes for differential
identification of the tissue(s) or cell type(s). For a number of disorders of the above
tissues or cells, particularly of the immune and neural system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to zinc finger protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of schizophrenia, kidney disease and other cancers. The tissue distribution in macrophage, breast, and kidney origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of tumors within these tissues, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 147

The translation product of this gene shares sequence homology with HNP36 protein, an equilibrative nucleoside transporter, which is thought to be important in gene transcription as well as serving as an important component of the nucleoside transport apparatus (See Accession No. 1845345). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTIICYLGLPRLEFYR
 YYQQLKLEGPGEQETKLDLISKGEEPRAGKEESGVSVSNSQPTNESHSIKAILK
 NISVLAFSVCFTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLG
 RSLTAVFMWPGKDSRWLPSWXLARLVFVPLLLLCNIKPRRYLTVVFEHDAWFI
 FFMAAFASFNGYLASLCMCFGPKKVKPAEAETAEPSPSSCVVWWHWGLFS
 PSCSGQLCDKGWTEGLPASLPVCLLPARGDPEWSGGFFF (SEQ ID
 NO:705); MSGQGLAGFFASVAMICAIASGSELSESAFGYFITACAVIILTIIC
 YLGLPRLEFYRYYQQLKLE GPGEQETKLDLISKGEEPRAGKEESGVSVSNSQ
 PTNESHSI (SEQ ID NO:706); SGVSVSNSQPTNESHSIKAILKNISVLAFSVCFI
 FTITIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLGRS (SEQ ID
 NO:707); TIGMFPAVTVEVKSSIAGSSTWERYFIPVSCFLT FNIFDWLGRSLTAVF
 MWPGKDSRWLPSWXLARLVFVPLLLLCNIK PRRYLTVVFEHDA (SEQ ID
 NO:708); FGPKKVKPAEAETAEPSPSSCVVWWHWGLFSPSCSGQLCDK

GWTEGLPASLPVCLLPLPSARGDPEWSGGFFF (SEQ ID NO:709). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in eosinophils and aortic endothelium and to a lesser extent in umbilical vein endothelial cell and thymus.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematopoietic disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
15 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to HNP36 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of blood neoplasias and other hematopoietic disease.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This gene is expressed primarily in breast cancer cell lines, thymus stromal cells, and ovary.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, endocrine and female reproductive system diseases including breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell
30 type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard
35 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of endocrine disorders. In addition, the tissue distribution in tumors of thymus, ovary, and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 149

Translation product of this gene has homology to pmt1 and pmt 2, two conserved schizosaccharomyces pombe genes. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

15 DDDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTFNEDEG
ELPEWVQVEEKQHRIRQLPVGKKEVEHYRKRWREINARPIXXXXXXXXXXXXX
XXXXXXXXLEQTRKKAEEVVNTVDIXRTRES (SEQ ID NO:710);
DDGFEIVPIEDPAKHRILDPEGLALGAVIASSKKAKRDLIDNSFNRYTF (SEQ
ID NO:711); KRWREINARPIXXXXXXXXXXXXXXXXXXXXXLEQTRKKAEE
AVVNTVDIXRTRES (SEQ ID NO:712). An additional embodiment is the
20 polynucleotide fragments encoding these polypeptide fragments (See Accession No.
e1216734).

This gene is expressed primarily in retina and ovary and to a lesser extent in breast cancer cell, epididymus and osteosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
25 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, neuronal growth disorders, cancer and reproductive system disorders.
Similarly, polypeptides and antibodies directed to these polypeptides are useful in
providing immunological probes for differential identification of the tissue(s) or cell
30 type(s). For a number of disorders of the above tissues or cells, particularly of the
neural and reproductive system, expression of this gene at significantly higher or lower
levels may be routinely detected in certain tissues (e.g., cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to
35 the standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those
comprising a sequence shown in SEQ ID NO: 382 as residues: Met-1 to Gly-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis or treatment of reproductive system disease and cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 150

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKEKKRNKKKKKTIGSPKRIQS
 PLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLS
 10 SLQSDPAGCVRPPAPNLAGAVEFNDVKTLREWITTISDPM EEDILQVVKYCTD
 LIEEKDLEKLDLVIKYMKRLMQQSVE SVWNMAFD FILDNVQVVLQQT YGSTLK
 VT (SEQ ID NO:713); MIKDKGRARTALTSSQPAHLCPENPLLHLKAAVKE
 KKRNNKKKTIGSPKRIQ (SEQ ID NO:714); KRIQSPLNNKLLNSPAKT
 LPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPAGCVRPP
 15 APNLAGAVEFNDVKTLREWITTISDPM (SEQ ID NO:715);
 TISDPM EEDILQVVKYCTDLIEEKDLEKLDLVIKYMKRLMQQSVE
 SVWNMAFD FILDNVQVVLQQT YGSTLKVT (SEQ ID NO:716). An additional
 embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in 12 week embryo and to a lesser extent in
 20 hemangiopericytoma and frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as
 reagents for differential identification of the tissue(s) or cell type(s) present in a
 biological sample and for diagnosis of diseases and conditions, which include, but are
 not limited to, growth disorders and hemangiopericytoma. Similarly, polypeptides and
 25 antibodies directed to these polypeptides are useful in providing immunological probes
 for differential identification of the tissue(s) or cell type(s). For a number of disorders
 of the above tissues or cells, particularly of the circular and neural system, expression
 of this gene at significantly higher or lower levels may be routinely detected in certain
 tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
 30 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an
 individual having such a disorder, relative to the standard gene expression level, i.e.,
 the expression level in healthy tissue or bodily fluid from an individual not having the
 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
 NO: 383 as residues: Leu-4 to Lys-11.

35 The tissue distribution indicates that polynucleotides and polypeptides
 corresponding to this gene are useful for the treatment of growth disorders,
 hemangiopericytoma and other soft tissue tumors.

FEATURES OF PROTEIN ENCODED BY GENE NO: 151

The translation product of this gene has been found to have homology to a human DNA mismatch repair protein PMS3. Preferred polypeptide fragments comprise the following amino acid sequence: FCHDCKFPEASPAMNCEP (SEQ ID NO:717). Also preferred are polynucleotide fragments encoding these polypeptide fragments (See Accession No. R95250).

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, lymphoma, immunodeficiency diseases, and cancers resulting from genetic instability. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 384 as residues: Met-1 to Lys-6.

The tissue distribution in neutrophils and the sequence homology indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of Hodgkin's lymphoma, since the elevated expression and secretion by the tumor mass may be indicative of tumors of this type. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Furthermore, its homology to a known DNA repair protein would suggest gene may be useful in establishing cancer predisposition and prevention in gene therapy applications.

FEATURES OF PROTEIN ENCODED BY GENE NO: 152

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, infectious diseases and lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of inflammation and infectious diseases.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 153

One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPFVLRKKC

NFFCWDSSAHSPLHPLSASCSAPACHASDTHLLYPSTRALCPSIFAWLVAPHS

20 VFRTNAPGPTPSSQSSPVFPVFPVSFMALIVCXLVCC (SEQ ID NO:720);

MASSVPAGGHTRAGGIFLIGKLDLEASLFKSFQWLPFVLRKKCNFFCWDSSAH

SLPLHPLSASCSAPACHA (SEQ ID NO:721);FAWL VAPHSVFRTNAPGPTPS

SQSSPVFPVFPVSFMALIVCXLVCC (SEQ ID NO:722). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammation and infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred

epitopes include those comprising a sequence shown in SEQ ID NO: 386 as residues: Ser-11 to Pro-17.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of infectious diseases and inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed in multiple tissues including ovary, uterus, adipose tissue, brain, and the liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, uterine, ovarian, brain, and liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic or therapeutic uses in the treatment of the female reproductive system, obesity, and liver disorders, particularly cancer in the above tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 155

This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3 (See Accession No. D87452).

This gene is expressed in multiple tissues including brain, aortic endothelial cells, smooth muscle, pituitary, testis, melanocytes, spleen, neutrophils, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological disorders including immunodeficiencies, cancers of the brain and the female reproductive system, as well as cardiovascular disorders, such as

atherosclerosis and stroke. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution suggest that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nervous system, including schizophrenia, neurodegeneration, neoplasia, brain cancer as well as cardiovascular and female reproductive disorders including cancer within the above tissues.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with the human gene encoding cytochrome b561 (See Accession No. P10897). Cytochrome b561 is a transmembrane electron transport protein that is specific to a subset of secretory vesicles

20 containing catecholamines and amidated peptides. This protein is thought to supply reducing equivalents to the intravesicular enzymes dopamine-beta-hydroxylase and alpha-peptide amidase. Preferred polypeptides of the invention comprise the amino acid sequence:

MAMEGYWRFLALLGSALLVGFLSVIFALVWVLHYREGLGWDGSALEFNWHP

25 VLMVTGFVFIQGLAIVYRLPWTWKCSKLLMKSIHAGLNAVAAILAISVVAVFE

NHNVNNIANMYSLHSWVGLIAVICYLLQLLSGFSVFLLPWAPLSLRAFLMPIHV

YSGIVIFGTVLATALMGLTEKLIFSLRDPAYSTFPPEGVFVNTLGLLLLVFGALIF

WIVTRPQWKRPKEPNSTILHPNGGTEQGARGSM PAYSGNNMDKSDSEL

30 NSEVAARKRNALDEAGQRSTM (SEQ ID NO:724); as well as antigenic fragments of at least 20 amino acids of this gene and/or biologically active fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system and metabolism related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product or RNA of this gene is useful for treatment or diagnosis of immune system and metabolic diseases or conditions including Tay-Sachs disease, phenylketonuria, galactosemia, various porphyrias, and Hurler's syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 157

The translation product of this gene shares sequence homology with collagen which is important in mammalian development. This gene also shows sequence homology with bcl-2. (See Accession No. P80988.) Preferred polypeptide fragments comprise the amino acid sequence: PGRAGPSPGLSLQLPAEPGHPAGNLAPLTSR PQPLCRIPAVPG (SEQ ID NO:725). Also preferred are polynucleotide sequences encoding this polypeptide fragment.

This gene is expressed primarily in HL-60 tissue culture cells and to a lesser extent in liver, breast, and uterus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunological diseases, hereditary disorders involving the MHC class of immune molecules, as well as developmental disorders and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those

comprising a sequence shown in SEQ ID NO: 390 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62.

The tissue distribution and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hereditary MHC disorders and particularly autoimmune disorders including rheumatoid arthritis, lupus, scleroderma, and dermatomyositis, as well as many reproductive disorders, including cancer of the uterus, and breast tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 158

This gene is expressed primarily in the amygdala region of the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, particularly those effecting mood and personality. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and/or diagnosis of a variety of brain disorders, particularly bipolar disorder, unipolar depression, and dementia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 159

This gene is expressed in a variety of tissues and cell types including brain, smooth muscle, kidney, salivary gland and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers of a variety of organs including brain, smooth muscle, kidney, salivary gland and T-cells and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the central nervous, urinary, salivary, digestive, and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
5 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain, smooth muscle, and T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of
10 various neurological, and cardiovascular disorders, but not limited to cancer within the above tissues. Additionally the gene product may be used as a target in the immunotherapy of the cancer. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma,
15 immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 160

The translation product of this gene shares sequence homology with collagen which is thought to be important in cellular interactions, extracellular matrix formation,
20 and has been found to be an identifying determinant in autoimmune disorders. Moreover, this gene shows sequence homology with the yeast protein, Sls1p, an endoplasmic reticulum component, involved in the protein translocation process in Yeast *Yarrowia lipolytica*. (See Accession No. 1052828; see also *J. Biol. Chem.* 271, 11668-11675 (1996).) With mouse, this same region shows sequence homology with
25 the heavy chain of kinesin. (See Accession No. 2062607.) Recently, suppression of the heavy chain of kinesin was shown to inhibit insulin secretion from primary cultures of mouse beta-cells. (See *Endocrinology* 138 (5), 1979-1987 (1997).) Moreover, kinesin was found associated with drug resistance and cell immortalization. (See 468355.) Thus, it is likely that this gene also act as a genetic suppressor elements.

30 This gene is expressed primarily in the greater omentum and to a lesser extent in a variety of organs and cell types including gall bladder, stromal bone marrow cells, lymph node, liver, testes, pituitary, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders of the endocrine, gastrointestinal, and immunological systems, including autoimmune disorders and cancers in a variety of organs and cell types.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and gastrointestinal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 393 as residues: Asn-27 to Leu-47, Gln-81 to Lys-88, Asp-93 to Lys-102, Asn-107 to Leu-116, Met-129 to Glu-141, Glu-150 to Asp-157, Lys-176 to Glu-185, Glu-333 to Tyr-349, Cys-393 to Leu-403, Gln-423 to Gly-429.

The tissue distribution in within various endocrine and immunological tissues combined with the sequence homology to a conserved collagen motif indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various autoimmune disorders including, but not limited to, rheumatoid arthritis, lupus erythematosus, scleroderma, dermatomyositis. Because the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene has homology to the tissue inhibitor of metalloproteinase 2. Such inhibitors are vital to proper regulation of metalloproteins such as collagenases (See Accession No. P16368). In addition, this gene maps to chromosome 17, and therefore, may be used as a marker in linkage analysis for chromosome 17 (See Accession No. P16368).

This gene is expressed primarily in several types of cancer including osteoclastoma, chondrosarcoma, and rhabdomyosarcoma and to a lesser extent in several non-malignant tissues including synovium, amygdala, testes, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, various types of cancer, particularly cancers of bone and cartilage, as well as various autoimmune disorders. Similarly, polypeptides and antibodies directed

to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various cancers and the sequence homology to a collagenase inhibitor indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 162

This gene is homologous to the mitochondrial ATP6 gene and therefore is likely a homolog of this gene family (See Accession No. X76197).

This gene is expressed primarily in brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, a variety of brain disorders, including Down's syndrome, depression, Schizophrenia, and epilepsy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain tissue indicates this gene is useful for diagnosis of various neurological disorders including, but not limited to, brain cancer.

Additionally the gene product may be used as a target in the immunotherapy of cancer in the brain as well as for the diagnosis of metabolic disorders such as obesity Tay-Sachs disease, phenylketonuria and Hurler's Syndrome.

FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in placenta, neutrophils, and microvascular endothelial cells and to a lesser extent in multiple tissues including brain, prostate, spleen, thymus, and bone.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neutropenia and other diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis various female reproductive disorders. Additionally the gene product may be used as a target in the immunotherapy of various cancers. Because the gene is expressed in some cells of lymphoid and endocrine origin, the natural gene product may be involved in immune functions and metabolism regulation, respectively. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 164

This gene is expressed primarily in neutrophils, monocytes, bone marrow, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune system disorders including, but not limited to, autoimmune disorders such as lupus, and immunodeficiency disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having
 5 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in various immune system tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of various immunological disorders such as Hodgkin's lymphoma, arthritis,
 10 asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 165

The translation product of this gene shares sequence homology with dystrophin which is thought to be defective in both Duchene and Becker Muscular Dystrophy.

- 15 Preferred polypeptide fragments comprise the following amino acid sequence:
 MKLLGECSSSIDSVKRLEHKLKEEEEESLPGFVNLHSTETQTAGVIDRWELLQAAQ
 ALSKELRMKQNLQKWQQFNSDLNSIWAWLGDTEEELEQLQRLELSTDIQTIELQ
 IKKLKELQKAVDHRKAILLSINLCSPEFTQADSKESRDLQDRLXQMNGRWDRV
 CSLLLEEWGRLQDALMQCQGFHEMSHGLLLMLENIDRRKNEIVPIDSNLDAEIL
 20 QDHHKQLMQIKHELLESQLRVASLQDMSCQLLVNAEGTDCLEAKEKVHVIGNR
 LKLLLKEVSRHIKELEKLLDVSSSQDLSSWSSADELDTSGSVSPXSGRSTPNR
 QKTPRGKCSLSQPGPSVSSPHSRSTKGGSDSSLSEXPGRSGRGFLFRVLRAA
 LPLQLLLLLLIGLACLVPMSSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
 NO:726); MKLLGECSSSIDSVKRLEHKLKEEEEESLPGFVNLHSTETQTAGVIDR
 25 WELLQAQALSKELRMKQNLQKWQQFNSDLNSIWAWLGDTEEELEQLQRLELS
 TDIQTIELQIK (SEQ ID NO:727); KLKELQKAVDHRKAILLSINLCSPEFTQADSK
 ESRDLQDRLXQMNGRWDRVCSLLLEEWGRLQDALMQCQGFHEMSHGLLLML
 ENIDRRKNEIVPIDSNLDAEILQDHHKQLMQIKHELLESQLRVASLQDMSCQL
 (SEQ ID NO:728); QDMSCQLLVNAEGTDCLEAKEKVHVIGNRLKLLLKEVS
 30 RHIKELEKLLDVSSSQDLSSWSSADELDTSGSVSPXSGRSTPNRQKTPRGKCS
 LSQPGPSVSSPHS (SEQ ID NO:729); DSSLSEXPGRSGRGFLFRVLRAAL
 PLQLLLLLLIGLACLVPMSSEEDYSCALSNNFARSFHPMLRYTNGPPPL (SEQ ID
 NO:730). Also preferred are polynucleotide fragments encoding these polypeptide
 fragments. Furthermore, this gene maps to chromosome 6, and therefore, may be used
 35 as a marker in linkage analysis for chromosome 6 (See Accession No. N62896).

This gene is expressed in numerous tissues including the heart, kidney, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, musculoskeletal disorders including Muscular Dystrophy and cardiovascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscle tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to dystrophin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of Muscular Dystrophy and other muscle disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the central nervous system, including Alzheimer's Disease, Parkinson's Disease, ALS, and mental illnesses. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 399 as residues: Pro-20 to Gly-26, Leu-37 to Pro-42, His-57 to Gly-63.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the central nervous system and may protect or

enhance survival of neuronal cells by slowing progression of neurodegenerative diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 167

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MKLLICGNYLAPSHSESSRRCCLLCFYPLCLEINFGMKVFLSMPFLVLFQ
SLIQED (SEQ ID NO:731). Polynucleotides encoding such polypeptides are also
provided. This gene is believed to reside on chromosome 15. Therefore polynucleotides
10 derived from this gene are useful in linkage analysis as chromosome 15 markers.

This gene is expressed primarily in human testes tumor and to a lesser extent in normal human testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the testes, particularly cancer, and other reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
20 the male reproductive tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
25 fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of testicular diseases including cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 168

30 This gene is expressed primarily in fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, conditions affecting hematopoietic development and metabolic diseases.
35 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the

hepatic system, and fetal hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 401 as residues: His-7 to Trp-17, Leu-19 to Lys-27, Pro-33 to Gly-44, Lys-68 to Gly-74, Lys-85 to Cys-95.

The tissue distribution indicates that the protein products of this gene are useful for treatment/diagnosis of diseases of the developing liver and hematopoietic system, and act as a growth differentiation factor for hematopoietic stem cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The polypeptide encoded by this gene is believed to be a membrane bound receptor. The extracellular domain of which is expected to consist of the following amino acid sequence:

RILLVKYSANEENKYDYLPTTFVNVCSELVKLVFCVLVSFCVIKKDHQSRNLKY
 ASWKEFSDFMKWSIPAFLYFLDNLIVFYVLSYLPAMAVIFS NFSIITALLFRIV
 LKXRLNWIQWASLLTLFLSIVALTAGTKTLQHNLAGRGFHHDAFFSPSNSCLL
 FRNECPRKDNCTAKEWTFPEAKWNTTARVFSHIRLGMGHVLIIVQCFISSMANI
 YNEKILKEGNQLTEXIFIQNSKLYFFGILFNGLTGLQRSNRDQIKNCGFFYGH
 S (SEQ ID NO:732). Thus, preferred polypeptides encoded by this gene comprise the extracellular domain as shown above. It will be recognized, however, that deletions of either end of the extracellular domain up to the first cysteine from the N-terminus and the first cysteine of the C-terminus, is expected to retain the biological functions of the full-length extracellular domain because the cysteines are thought to be responsible for providing secondary structure to the molecule. Thus, deletions of one or more amino acids from either end (or both ends) of the extracellular domain are contemplated. Of course, further deletions including the cysteines are also contemplated as useful as such polypeptides is expected to have immunological properties such as the ability to evoke and immune response. Polynucleotides encoding all of the foregoing polypeptides are provided.

This gene is expressed primarily in human osteoclastoma and to a lesser extent in hippocampus and chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, cancers, particularly those of the bone and connective tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 402 as residues: Met-1 to Cys-6, Ala-41 to Tyr-49, Lys-76 to Lys-84.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of cancers of the bone and connective tissues, and may act as growth factors for cells involved in bone or connective tissue growth.

FEATURES OF PROTEIN ENCODED BY GENE NO: 170

Preferred polypeptides encoded by this gene comprising the following amino acid sequence:

NSVPNLQTLAVLTEAIGPEPAIPRXPREPPVATSTPATPSAGPQPLPTGTV

LVPGGPAPPCLGEAWALLPPCRPSLTSCFWSPRPSPWKETGV (SEQ ID NO:733). Polynucleotides encoding such polypeptides are also provided herein.

This gene is expressed primarily in hematopoietic progenitor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the blood including cancer and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the blood/circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 403 as residues: Gln-4 to His-10, Pro-25 to His-32.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis of diseases involving growth differentiation of hematopoietic cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 171

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequences: ALQLAFYPDAVEEWLEENVHPSLQRLQXLLQDLSEVSAPP (SEQ ID NO:734); and/or CHPPALAGTLLRTPEGRAHARGLLLEAGGA (SEQ ID NO:735). Polynucleotides encoding such polypeptides are also provided. The protein product of this gene shares sequence homology with metallothionines. Thus, polypeptide encoded
10 by this gene are expected to have metallothionine activity, such activities are known in the art and described elsewhere herein.

This gene is expressed primarily in kidney cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases of the kidney including cancer and renal dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system,
20 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder; relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the
25 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 404 as residues: Ser-47 to Gln-52.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of diseases of the kidney including kidney failure.
30

FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
35 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 405 as residues: Gln-31 to Thr-43, Gly-51 to Ser-58, Pro-65 to Pro-72.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of developmental problems with fetal tissue. The gene may be involved in vital organ development in the early stage, especially hematopoiesis, cardiovascular system, and neural development.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 173

The translation product of this gene shares sequence homology with TGN38, an integral membrane protein previously shown to be predominantly localized to the trans-Golgi network (TGN) of cells.

This gene is expressed primarily in developing embryo and to a lesser extent in cancer tissues including lymphoma, endometrial, prostate and colon.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 406 as residues: His-65 to Ser-72, Pro-82 to Gly-91, Pro-98 to Glu-118, Ser-126 to Gly-166, Pro-180 to Asp-188, Tyr-209 to Lys-214, Gln-220 to Leu-228.

The tissue distribution and homology to an integral membrane protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for

diagnosis of cancers and developmental abnormalities where aberrant expression relates to an abnormality.

FEATURES OF PROTEIN ENCODED BY GENE NO: 174

5 The translation product of this gene shares sequence homology with a dnaJ heat shock protein from E. coli which is allelic to sec63, a gene that affects transit of nascent secretory proteins across the endoplasmic reticulum in yeast.

This gene is expressed primarily in Hodgkin's lymphoma and to a lesser extent in testes.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification
15 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to
20 the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 407 as residues: Thr-13 to Trp-21, Arg-74 to Asp-81.

25 The tissue distribution and homology to dnaJ indicates that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic for cancer including Hodgkin's lymphoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 175

30 This gene is expressed primarily in endothelial cells and to a lesser extent in bone marrow stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, diseases involving angiogenic abnormalities including diabetic
35 retinopathy, macular degeneration, and other diseases including arteriosclerosis and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for treating diseases where an increase or decrease in angiogenesis is indicated and as a factor in the wound healing process.

FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with MAT8 (mouse) which is thought to be important in regulating chloride conductance in cells (particularly in the breast) by modulating the response mediated by cAMP and protein kinase C to extracellular signals.

This gene is expressed primarily in amniotic cells and hematopoietic cells including macrophages, Neutrophils, T cells, TNF induced aortic endothelium and to a lesser extent in testes, TNF induced epithelial cells, and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, inflammatory responses mediated by T cells, macrophages, and/or neutrophils particularly those involving TNF, and also cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 409 as residues: Thr-19 to Ala-33, Leu-54 to Asp-82, Pro-89 to Ala-97, Pro-100 to Lys-125, Ser-127 to Phe-135, Gly-164 to Leu-169, Cys-173 to Arg-178.

The tissue distribution and homology to mat-8 indicates that polynucleotides and polypeptides corresponding to this gene are useful for modifying inflammatory

responses to cytokines such as TNF and thus modifying the duration and/or severity of inflammation. Polynucleotides and polypeptides derived from this gene are thought to be useful in the diagnosis and treatment of cancer.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 177

This gene is expressed primarily in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vascular restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases associated with vascular response to injury such as vascular restenosis following angioplasty..

FEATURES OF PROTEIN ENCODED BY GENE NO: 178

One embodiment of the claimed invention comprises:

25 MRPDWKAGAGPGGPPQKPAPSSQRKPPARPSAAAAAIAVAAAEERRLRQRN
RLRLEEDKPAVERCLEELVFGDVENDEALLRRLRGPRVQEHEDSGDSEVENEAKGNFPPQKKPVWVDEEDEDEEMVDMMNRRFRKDMMKNAESKLSKDNLKKRLKEEFQHAMGGVPAWAETTKRKTSSDDESEEDDDLLQRTGNFISTSTSLPRGILKMKNQCQHANAEPTVARISICAVPSRCTDCDGCWD (SEQ ID NO:737); or
30 CLEELVFGDVENDEALLRRLRGPRVQEHEDSGDSEVENEAKGNFPPQKKPVWVDEEDEDEEMVDMMNRRFRKDMMKNAESKLSKDNLKKRLKEEFQHAMGGVPAWAETTKRKTSSDDESEEDDDLLQRTGNFISTSTSLPRGILKMKNQCQHANAEPTVARISICAVPSRCTDCDGC (SEQ ID NO: 738). LKEKIVRSFEVSPDGS
FLLINGIAGYLHLLAMKTKELIGSMKINGRVAASTFSSDSKKVYASSGDGEVYV
35 WDVNSRKCLNRFVDEGSLYGLSIATSRNGQYVACGSNCGVVNIYNQDSCLQE
TNPKPIKAIMNLVTGVTSLTFNPTTEILAIASEKMKEAVRLVHLPSCVTF\$NFPVI
KNKNISHVHTMDFSPRSGYFALGNEKGKALMYRLHHYSDF (SEQ ID NO:739);

and/or KINGRVAASTFSSDSKKVYASSGDGEVYVWDVNSRKCLNRFVDEGSL
YGLSLATSRNGQYVACGSNCGVVNIYNQDSCLQETNPKPIKAIMNLVTGVTSLT
FNPTTEILAIASEKMKEAVRLVHLPSCTVFSNFPVIKNKNISHVHTMDFSPRSG
YFALGNEKGKAL (SEQ ID NO:740).

5 This gene is expressed primarily in epididymus and endometrial tumors and to a lesser extent in T cell lymphoma and cell lines derived from colon cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, tumors of the reproductive organs including testis and endometrial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 411 as residues: Ser-67 to Lys-72, Val-87 to Leu-93, Tyr-128 to Pro-141, Asp-204 to Gly-210.

The tissue distribution indicates that the protein products of this gene are useful for treating tumors of the endometrium or epithelial tumors of the reproductive system.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 179

Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MRILQLILLALATGLVGGETRIIKGFECKLHSQPWQAALFEKTRLLCGATLIAPR
WLLTAAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNNSLPNKDH
30 RNDIMLVKMASPVSITWAVRPLTLSSRCVTAGTSCSFPAGAARPDPSYACLTPC
DAPTSPSLSTRSVRTPTPATSQTPWCVPACRKGARTPARVTPGALWSVTSLFKA
LSPGARIRVRSPESLVSTRKSANMWTGSRRR (SEQ ID NO:741); ETRIIKGFEC
KLHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQHNLQKEE
GCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPVSITWAVRPLTLSSR
35 CVTAGTSCSFPAGAARPDPSYACLTPCDAPTSPSLSTRSVRTPTPATSQTPWCVP
ACRKGARTPARVTPGALWSVTSLFKALSPGARIRVRSPESLVSTRKSANMWTG

SRRR (SEQ ID NO:742); or CKLHSQPWQAALFEKTRLLCGATLIAPRWLLT
AAHCLKPRYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNS

(SEQ ID NO:743). The translation product of this gene shares sequence homology
with neuropsin a novel serine protease which is thought to be important in modulating
extracellular signaling pathways in the brain. Owing to the structural similarity to other
serine proteases the protein products of this gene are expected to have serine protease
activity which may be assayed by methods known in the art and described elsewhere
herein.

This gene is expressed primarily in endometrial tumor and to a lesser extent in
colon cancer, benign hypertrophic prostate, and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, cancers of the endometrium or colon and benign hypertrophy of the
prostate. Similarly, polypeptides and antibodies directed to these polypeptides are
useful in providing immunological probes for differential identification of the tissue(s)
or cell type(s). For a number of disorders of the above tissues or cells, particularly of
the urogenital or reproductive systems, expression of this gene at significantly higher or
lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to
the standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those
comprising a sequence shown in SEQ ID NO: 412 as residues: Gly-12 to Ser-22, Pro-
34 to Ser-53.

The tissue distribution and homology to serine proteases indicates that
polynucleotides and polypeptides corresponding to this gene are useful for diagnosing
or treating hyperproliferative disorders such as cancer of the endometrium or colon and
hyperplasia of the prostate.

FEATURES OF PROTEIN ENCODED BY GENE NO: 180

Preferred polypeptide encoded by this gene comprise the following amino acid
sequence: VLQGRYFSPILEMRRLRPEGXXNLPGGSRAQKEPRQDLTLVLWPHC
PHFAMTRSYPVKQCMVQGSFYCFIFKGPVQNW (SEQ ID NO:744).

Polynucleotides encoding such polypeptide are also provided.

This gene is expressed primarily in fetal brain

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, identifying and expanding stem cells in the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for detecting and expanding stem cell populations in the (or of the) central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in early stage human brain and a stromal cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, developmental abnormalities of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 414 as residues: Gln-42 to Gln-47, Gln-54 to Pro-60.

The tissue distribution indicates that the protein products of this gene play a role in the development of the central nervous system. Therefore this gene and its products

are useful for diagnosing or treating developmental abnormalities of the central nervous system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 182

5 Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MPIIDQVNPELHDFMQSAEVTIFALSWLITWFGHVLSDFRHVVRLYDF
FLACHPLMPIYFAAVIVLYREQEVLDCDCDMASVHHLLSQIPQDLPYETLISRXE
TFLFSFHPNLLGRPLPNSKLRGRQPLLSKTLSWHQPSRGLIWCCGSGXRGLL
10 RPEDRTKDVLTTPRTNRFVKLAVMGLTVALGAAALAVVKSALWAPKFQLQL
FP (SEQ ID NO:745); or CPEFFIPATLPCPFVFAFTSEASSRAYLTQRGPGGLAQ
NLMPLPVGFWMGSLPPPWCWRKWVSEACSCFC (SEQ ID NO:746) These
polypeptides are structurally similar to various TGF-beta family members. Thus, this
polypeptide is expected to have a variety of activities in the modulation of cell growth
15 and proliferation.

This gene is expressed primarily in osteoclastoma, microvascular endothelium, and bone marrow derived cell lines.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
20 biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, hematological diseases particularly involving aberrant proliferation of stem cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
25 the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 415 as residues: Ser-33 to Ala-39.

The tissue distribution indicates that the protein products of this gene is useful for treating disorders of the progenitors of the immune system. Applications include in vivo expansion of progenitor cells, ex vivo expansion of progenitor cells, or the
35 treatment of tumors of the circulatory system, such as lymphomas.

FEATURES OF PROTEIN ENCODED BY GENE NO: 183

This gene maps to chromosome 17 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 17. In specific embodiments, polypeptides of the invention comprise the sequence:

- 5 GFGSVSA.AGRRSGGTWQPVQ (SEQ ID NO:747); PGGLAVGSRWWSRSLT (SEQ ID NO:748); LEPSRQRRPRRRGGTSRPETDQRAKCWRQL (SEQ ID NO:749); and/or VCLRCQNRMEN (SEQ ID NO:750). In further specific embodiments, polypeptides of the invention comprise the sequence: MAACTARRPGR GQPLVVPVADXGPVAKAALCAAXAGAFSPASTTTTTRHLSSNRNRPPEGKVLETV
- 10 GVFEVPKQNGKYETGQLFLHSIFGYRGVVLFPWQARLXDRDVASAAPEKAEN PAGHGSKEVKGKTHTYYQVLIDARDCPHISQRSQTEAVTFLANHDDSRALYAIP GLDYVSHEDILPYTSTDQVPIQHELFFERFLLYDQTKAPPFVARETLRAWQEKNH PWLELSDVHRETTENIRVTVIPFYMGMREAQNSHVYWWRYCIRLENLDSDVVQ LRERHWRIFSLSGTLETVRGRGVVGREPVLSKEQPAFQYSSHVSLQASSGHMW
- 15 GTFRFERPDGSHFDVRIPFSLESNKDEKTPPSGLHW (SEQ ID NO:751); MAACTARRPGRGQPLVVPVADXGPVAKAALCAA (SEQ ID NO:752); VLETVG VFEVPKQNGKYETGQLFLHSIFGYRGVVL (SEQ ID NO:757); GLDYVSHEDILPYTST (SEQ ID NO:758); DVHRETTENIRVTVIPFYM (SEQ ID NO:759); WWRYCIRLENLDSDVVQLRER (SEQ ID NO:760); and/or PAFQYSS
- 20 HVSLQASSGHMWGTFRFER (SEQ ID NO:761). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in gall bladder, prostate, and fetal brain, and to a lesser extent in a few tumor and fetal tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, growth related disorders such as cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
- 30 of the above tissues or cells, particularly of the prostate, gall bladder, and fetal brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,
- 35 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of growth-related disorders, such cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 184

In specific embodiments, polypeptides of the invention comprise the sequence: SLCCPEGAEGC (SEQ ID NO: 762) and/or QLKKTHYDRPCP (SEQ ID NO: 763). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in stromal cell, tonsil, and glioblastoma and to a lesser extent in some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune and inflammatory disorders and glioblastoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, tonsil, and glioblastoma expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Additionally, it is believed that the product of this gene regulates pancreatic cell differentiation into beta cells. Accordingly, polynucleotides and polypeptides of the invention are useful in the treatment of insulin-dependent diabetes mellitus and associated conditions e.g. pancreatic hypofunction and the prevention, as well as the treatment of undifferentiated type pancreatic cancers. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 417 as residues: Pro-27 to Ala-32.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune and inflammatory disorders and glioblastoma.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 185

This gene is expressed primarily in hepatocellular carcinoma and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 418 as residues: Gly-32 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in hippocampus and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 187

This gene is expressed primarily in bone cancer and hippocampus and to a lesser extent in osteoclastoma and other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, bone-related disorders and neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, osteoclast, and hippocampus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of bone-related disorders and neuronal diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 188.

This gene maps to chromosome 4 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 4.

This gene is expressed primarily in neuronal tissues such as hippocampus, spinal cord, and hypothalamus and to a lesser extent in a few other tissues such as ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 189

This gene maps to chromosome 10, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 10.

5 This gene is expressed primarily in neuronal tissues and immune tissues, and to a lesser extent in a few other tissues such as skin tumor, lung etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neuronal and immune-related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuronal and immune-related tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 422 as residues: Pro-19 to Asp-25.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neuronal and immune-related disorders.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 190

The translation product of this gene shares sequence homology with human N33, a gene located in a homozygously deleted region of human metastatic prostate cancer which is thought to be important in prevention of prostate cancer. In specific embodiments, polypeptides of the invention comprise the sequence:

30 AQRKKEMVLSEKVSQLEWTKRNPVIRMGDKFRRLVKAPPRNYSVIVMFTALQLHRQCVVCKQADEEFQLANSWRYSSAFTNRIFFAMVDFDEGSDVFQMLNMNSAPTFINFPKKGPKRGDTYELQVRGFSAEQIARWLADRTDVNIRVIRPPNMAARWRFWCVSVT (SEQ ID NO:765); MVVALLIVCDVPSAS (SEQ ID NO:766); AQRKKEMVLSEKVSQLEWTKRNPVIRMGDKF (SEQ ID:768); RRLVKAPPRNYSVIVMFTALQLHRQCVVCKQADEEFQLANSWRYSSAFTNRIFFA (SEQ ID NO:769); MVDVDFDEGSDVFQMLNMNSAPTFINFPKKGPK (SEQ ID NO:770); KRGDTYELQVRGFSAEQIARWLADRTDVNIRVIRPPN

(SEQ ID NO:771); and/or YAGPLMLGLLLA VIGGLVYLRRVTWNFSLIKLDGLLQL CVLCLL (SEQ ID NO:772). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in infant adrenal gland prostate cell line and to a lesser extent in a few other tissues like liver, smooth muscle etc.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, prostate cancer and endocrine disorders. Similarly, polypeptides and
10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate and adrenal gland, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
15 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 423 as residues: Pro-34 to Gly-43, Arg-113 to Pro-120.

20 The tissue distribution and homology to N33 indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment for prostate cancer and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 191

25 This gene is expressed primarily in T cell and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to
30 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal
35 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue

or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 424 as residues: Trp-3 to Phe-9.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 192

This gene maps to chromosome 6, therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 6. Neural activity and neurotrophins induce synaptic remodeling in part by altering gene expression. This gene is believed to be a glycosylphosphatidylinositol-anchored protein encoded by a hippocampal gene and to possess neural activity. This molecule is believed to be expressed in postmitotic-differentiating neurons of the developing nervous system and neuronal structures associated with plasticity in the adult. Message of this gene is believed to be induced by neuronal activity and by the activity-regulated neurotrophins BDNF and NT-3. The product of this gene is believed to stimulate neurite outgrowth and arborization in primary embryonic hippocampal and cortical cultures and to act as a downstream effector of activity-induced neurite outgrowth. In specific embodiments, polypeptides of the invention comprise the sequence: DAVFKGFSDCLLKLGD (SEQ ID NO:773); CQEGAKDMWDKLRKESKNLN (SEQ ID NO:774); VLLVLSAALATWLSF (SEQ ID NO:775); MGLKLNGRYISLILAVQIAYLVQAVR AAGKCD (SEQ ID NO:776); PAAWDDKTNIKTVCTYW EDFHSCTVTALTDCQEGAKDMWDKLRKESKNLN (SEQ ID NO:777); and/or MGLKLNGRYISLILAVQIAYLVQAVRAAGKCD (SEQ ID NO:778). Polynucleotides encoding this polypeptide are also encompassed by the invention.

This gene is expressed primarily in human-placenta, endometrial tumor and tissues of the central nervous system (CNS).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, relating to reproductive disorders, cancers and neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and neurological disorders, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 425 as residues: Asp-47 to Asp-63, His-75 to Tyr-80, Pro-83 to Tyr-89.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive disorders such as endometrial tumors. Expression of this gene in tissues of the CNS and its strong homology to Neuritin suggest that the protein product from this gene may also be used in the treatment and diagnosis of neurological disorders and in the regeneration of neural tissues, e.g., following injury.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 193

The translation product of this gene shares sequence homology with tenascin which is thought to be important in development. The translation product of this gene is believed to be a ligand of the fibroblast growth factor family. FGF ligand activity is known in the art and can be assayed by methods known in the art and disclosed elsewhere herein.

This gene is expressed primarily in endometrial tumors, and other types of tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 426 as residues: Gly-29 to Glu-34, Arg-71 to Arg-76, Thr-176 to Cys-182, Gly-184 to Glu-199.

The tissue distribution and homology to tenascin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 194

In specific embodiments, polypeptides of the invention comprise the sequence: MNSAAGFSHLDRRERVVLKLGESFEKQPRCASTLC (SEQ ID NO:779).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal human lung and neutrophils.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to: lung development and respiratory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
20 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal lung and neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis
25 and treatment of lung and immunity related diseases, for example, lung cancer, viral, fungal or bacterial infections (e.g. lesions caused by tuberculosis), inflammation (e.g. pneumonia), metabolic lesions etc.

FEATURES OF PROTEIN ENCODED BY GENE NO: 195

30 This gene is expressed primarily in breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immunal disorders. Similarly, polypeptides and antibodies directed to
35 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immunal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 196

This gene maps to chromosome 5 and accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 5. The translation product of this gene shares sequence homology with human M-phase phosphoprotein 4 which is thought to be important in phosphorylation and signal transduction processes. In

specific embodiments, polypeptides of the invention comprise the sequence:

TIYPTEEELQAVQKIVSITERALKLVSD (SEQ ID NO:780); RALKGVLRV
GVLAKGLLLLRGDRNVNLVLLC (SEQ ID NO:781); ALAALRHAKWFQARAN
GLQSCVIRILRDLCQRVPTWS (SEQ ID NO:782); GDALRRVFECISSGIL (SEQ
ID NO:783); LAFRQIHKVLGMDPLP (SEQ ID NO:784); and/or TIYPTEEELQAVQ
KIVSITERALKLVSDSLSEHEKNKNKEGDDKKEGGKDRALKGVLRVGVLAGK
LLLLRGDRNVNLVLLCSEKPSKTLISRIAENLPKQLAVISPEKYDIKCAVSEAAIIL
NSCVEPKMQVTITLTSPIREENMREGDVTSGMVKDPPDVLDRQKCLDALAALR
HAKWFQARANGLOSCVIRILRDLCQRVPTWSDFPSWAMELLVEKAISSASSP
QSPGDALRRVFECISSGILKGSPGLLDPCCKDPFDTLATMTDQQREDITSSAQFA
LRLLAFRQIHKVLGMDPLPQMSQRFNIHNNRKRRRDSDGVDGFEGGKKDKK
DYDNF (SEQ ID NO:785). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Hippocampus and to a lesser extent in Prostate, Human Frontal Cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, disorders related to reproductive system and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system and nervous system, expression of this gene at significantly higher or lower

levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human M-phase phosphoprotein 4 indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and nervous system disorders.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 197

In specific embodiments, polypeptides of the invention comprise the sequence: MGSQHSAAARPSSCRRKQEDDRDG (SEQ ID NO:786); LLAEREQEEAIAQFPYVEFTGRDSITCLTC (SEQ ID NO:787); and/or QGTGYIPTEQVNELVALIPHSDQRLRPQRTKQYV (SEQ ID NO:788).

15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Human Primary Breast Cancer and to a lesser extent in Human Adult Spleen, Hodgkin's Lymphoma I, Salivary Gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, cancer and immunal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 430 as residues: Ser-126 to Gly-138.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and immunal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 198

This gene is expressed primarily in monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, blood cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of blood cell disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 199

This gene is expressed primarily in Human Ovary and Synovia and to a lesser extent in Human 8 Week Whole Embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, reproductive and developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and developmental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of reproductive and developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 200

This gene maps to chromosome 8 and therefore polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 8. The translation product of this gene shares limited sequence homology with collagen proline rich domain.

This gene is expressed primarily in CNS.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential-identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 433 as residues: Pro-35 to Asp-41.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 201

Translation product of this gene shares homology with a mammalian histone H1a protein. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: ARLNVGRESLKREMLKSQGVKVSESPMGAR HSSWPEGAAAFCKKVQGAQMFPFRR (SEQ ID NO:789); ARLNVGRESLKR EML (SEQ ID NO:790); LKSQGVKVSESPMGARHSSW (SEQ ID NO:791); AAFCKKVQGAQMFPFRR (SEQ ID NO:792). An additional embodiment is the polynucleotide fragments encoding these polypeptide (See Accession No. pirlS24178) fragments.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are

not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at

5 significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in vital immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such

15 as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 202

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as

20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above

25 tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

30 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for

35 immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.



FEATURES OF PROTEIN ENCODED BY GENE NO: 203

This gene is expressed primarily in Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
5 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, infectious disorders, immune disorders, and cancers. Similarly,
polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For
10 a number of disorders of the above tissues or cells, particularly of the immune system,
expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
15 the expression level in healthy tissue or bodily fluid from an individual not having the
disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID
NO: 436 as residues: Thr-31 to Lys-36.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for diagnosis and treatment of infectious
20 disorders, immune disorders, and cancers. Since the gene is expressed in cells of
lymphoid origin, the natural gene product may be involved in immune functions.
Therefore it may be also used as an agent for immunological disorders including
arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as
well as, antibodies directed against the protein may show utility as a tumor marker
25 and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 204

This gene maps to chromosome 16 and therefore polynucleotides of the
invention can be used in linkage analysis as markers for chromosome 16. The
30 translation product of this gene shares sequence homology with lactate dehydrogenase
which is thought to be important in lactate metabolism.

This gene is expressed primarily in human tonsils and to a lesser extent in
Spleen, and Neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as
35 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, immune disorders, infectious disorders, and cancers. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune disorders, infectious disorders, and cancers, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 437 as residues: Gly-7 to Ser-12.

The tissue distribution and homology to lactate dehydrogenase gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders, infectious disorders, and cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 205

The translation product of this gene shares sequence homology with Gcap1 protein which is developmentally regulated in brain.

This gene is expressed primarily in placenta and endometrial tumor and to a lesser extent in several other tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions, which include, but are not limited to, vasculogenesis/angiogenesis and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to Gcap1 protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorder or dysfunction of vascular system of tumorigenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 206

In specific embodiments, polypeptides of the invention comprise the sequence
MPYAQWLAENDRFEEAQKAFHKAGRQREA (SEQ ID NO:799);
VQVLEQLTNNAVAESRFNDAAYYWMLSMQCLDLAQD (SEQ ID NO:794);
5 PAQKDTMLGKFYHFQRLAELYHGYHAIHRHTEDP (SEQ ID NO: 795);
FSVHRPETLFNISRFLHSLPKDTPSGISKVKILFT (SEQ ID NO:800);
LAKQSKALGAYRLARHAYDKLRGLYIP (SEQ ID NO:796); ARFQKSIELG
TLTIRAKPFHDSEELVPLCYRCSTNN (SEQ ID NO: 797); and/or PLLNNLGNVC
INCRQPFIFSASSYDVLHLVEFYLEEGITDEEAISLIDLEVLRPKRDDRQLEICKQQ
10 LPDSCG (SEQ ID NO:798). Polynucleotides encoding these polypeptides are also
encompassed by the invention.

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
15 biological sample and for diagnosis of diseases and conditions, which include, but are
not limited to, male reproductive and endocrine disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the reproductive and endocrine systems,
20 expression of this gene at significantly higher or lower levels may be routinely detected
in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum,
plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
25 disorder.

The tissue distribution indicates that polynucleotides and polypeptides
corresponding to this gene are useful for treatment of male reproductive and endocrine
disorders.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 207

This gene is expressed in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions, which include, but are
35 not limited to, lung diseases such as cystic fibrosis. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the respiratory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
5 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 440 as residues: Tyr-49 to Cys-54.

The tissue distribution indicates that polynucleotides and polypeptides
10 corresponding to this gene are useful for detection and treatment of disorders associated with developing lungs particularly in premature infants where the lungs are the last tissues to develop. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of lung tumors since the gene may be involved in the regulation of cell division,
15 particularly since it is expressed in fetal tissue. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO.	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO.	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HLHDS67	97979 03/27/97	Uni-ZAP XR	11	2526	427	2526	458	458	234	1	30	31	30
2	HLHDZ58	97979 03/27/97	Uni-ZAP XR	12	1131	1	1131	129	129	235	1	14	15	115
3	HLMMJ13	97979 03/27/97	Lambda ZAP II	13	941	39	941	62	62	236	1	44	45	102
3	HLMMJ13	97979 03/27/97	Lambda ZAP II	218	941	39	941	245	245	441	1	35	36	41
4	HLTE125	97979 03/27/97	Uni-ZAP XR	14	843	1	843	155	155	237	1	19	20	42
5	HMSJX24	97979 03/27/97	Uni-ZAP XR	15	1018	1	1018	90	90	238	1	18	19	36
6	HNFED65	97979 03/27/97	Uni-ZAP XR	16	661	1	661	76	76	239	1	28	29	127
7	HNHDX07	97979 03/27/97	Uni-ZAP XR	17	553	1	553	106	106	240	1	23	24	66
8	HNHGC82	97979 03/27/97	Uni-ZAP XR	18	869	1	869	101	101	241	1	21	22	68
9	HNHGO09	97979 03/27/97	Uni-ZAP XR	19	959	1	959	176	176	242	1	21	22	44
10	HOUBE18	97979 03/27/97	Uni-ZAP XR	20	1446	1	1446	101	101	243	1	27	28	50
11	HOUDL69	97979 03/27/97	Uni-ZAP XR	21	1471	579	1460	692	692	244	1	31	32	42

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
12	HPMF171	97979 03/27/97	Uni-ZAP XR	22	1402	242	1402	401	245	1	32	33	60
13	HPMGQ55	97979 03/27/97	Uni-ZAP XR	23	1047	1	1047	164	246	1	26	27	35
14	HPQAC69	97979 03/27/97	Lambda ZAP II	24	990	1	988	82	247	1	20	21	37
15	HP1BB03	97979 03/27/97	Uni-ZAP XR	25	1208	350	1173	398	248	1	29	30	210
16	HP1WA66	97979 03/27/97	pBluescript	26	1922	1381	1922	24	249	1	33	34	547
16	HP1WA66	97979 03/27/97	pBluescript	219	575	1	575	148	442	1	22	23	65
17	HP1WC08	97979 03/27/97	pBluescript	27	1951	1422	1874	219	250	1	19	20	299
18	HRGCZ46	97979 03/27/97	Uni-ZAP XR	28	3989	2635	3989	2748	251	1	16	17	39
19	HSADV34	97979 03/27/97	Uni-ZAP XR	29	3735	2966	3735	272	252	1	30	31	594
19	HSADV34	97979 03/27/97	Uni-ZAP XR	220	3018	1929	3018	26	443	1	1	2	156
20	HSDFW61	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	30	1667	59	1625	138	253	1	32	33	130
21	HSDFP60	97974 04/04/97	Uni-ZAP XR	31	1408	1	1408	285	254	1			20

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
22	HSOA155	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	32	2031	1273	2031	1285	1285	255	1	29	30	30
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	33	971	13	971	91	91	256	1	19	20	218
23	HSQEO84	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	221	968	8	968	86	86	444	1	20	21	56
24	HISXAM05	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	34	1792	369	1792	470	470	257	1	26	27	49
25	HISXAS67	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	35	896	1	896	96	96	258	1	32	33	121
26	HTIDAF28	97974 04/04/97 209080 05/29/97	pSport1	36	912	1	912	38	38	259	1	22	23	87
27	HTEGQ64	97974	Uni-ZAP XR	37	1382	67	1382	271	271	260	1			25

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO. X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO. Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/04/97 209080 05/29/97											
28	HTGEU09	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	38	872	1	872	74	261	1	18	19	28
29	HTOAM21	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	39	812	1	812	41	262	1	30	31	43
30	HTPBW79	209511 12/03/97	Uni-ZAP XR	40	1515	118	1507	302	263	1	24	25	362
30	HTSEV09	97974 04/04/97 209080 05/29/97	pBluescript	222	1404	1	1265	92	445	1	19	20	415
31	HJPCD40	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	41	704	22	704	117	264	1	18	19	127
32	HTWBY48	97974 04/04/97 209080 05/29/97	pSport1	42	1094	1	1094	32	265	1	34	35	53
33	HTWC146	97974 04/04/97	pSport1	43	1821	892	1647	56	266	1	26	27	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209080 05/29/97												
34	HITXG175	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	44	1024	30	1024		167	267	1	20	21	25
35	HWTFB59	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	45	983	779	983	85	85	268	1	30	31	221
35	HWTFB59	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	223	707	488	707	514	514	446	1	41	42	64
36	HADAIE74	97974 04/04/97 209080 05/29/97	pSport1	46	2421	664	1587	710	710	269	1			2
37	HACIFB60	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	47	840	1	840	97	97	270	1	30	31	48
38	HATF60	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	48	2432	1193	2246	1491	1491	271	1	17	18	51
39	HBMSN25	97974	Uni-ZAP XR	49	1742	1165	1742	1207	1207	272	1	23	24	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
40	HCDAR68	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	50	1487	181	1455	325	325	273	1	35	36	56
41	HCE3179	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	51	1328	251	1328	525	525	274	1			21
42	HMDAN54	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	52	1856	725	1853	928	928	275	1	33	34	50
43	HCECA49	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	53	1558	310	1408	393	393	276	1			1
44	HCEEC15	97974 04/04/97 209080 05/29/97	Uni-ZAP XR	54	948	1	948	9	9	277	1	23	24	65
45	HCESE40	97974 04/04/97 209080 05/29/97	pBluescript	55	990	99	990	193	193	278	1	32	33	256

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
45	HCESEF40	97974 04/04/97 209080 05/29/97	pBluescript	224	1384	99	1384	193	193	447	1	32	33	205
46	HCFMV39	97974 04/04/97 209080 05/29/97	pSport1	56	1603	1	1296	96	96	279	1	29	30	102
47	HCMSSX86	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	57	1052	5	786	12	12	280	1	28	29	32
48	HCNAP62	97975 04/04/97 209081 05/29/97	Lambda ZAP II	58	814	1	558	93	93	281	1	22	23	42
49	HICRAF32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	59	1215	257	1215		356	282	1	19	20	20
50	HCUDC07	97975 04/04/97 209081 05/29/97	ZAP Express	60	478	1	478	147	147	283	1	36	37	69
51	HCWBB42	97975 04/04/97 209081	ZAP Express	61	618	1	618	212	212	284	1	35	36	74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of First Signal Pep	5' NT of AA of First Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
52	HDTAB05	97975 04/04/97 209081 05/29/97	pcMVSpot 2.0	62	751	1	751	257	257	285	1	21	22	32
53	HE2AV74	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	63	780	283	780		433	286	1			16
54	HE2AV71	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	64	588	21	588	169	169	287	1			16
55	HE2GS36	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	65	774	272	774	445	445	288	1			37
56	HE2OF09	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	66	1866	1313	1866	1596	1596	289	1			11
57	HE6EU50	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	67	1152	117	686	237	237	290	1	20	21	34
58	HE9HU17	97975 04/04/97	Uni-ZAP XR	68	2483	1577	2448	1620	1620	291	1			14

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
59	HE9ND48	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	69	536	1	536	83	83	292	1	36	37	43
60	HEBBW11	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	70	865	647	865		388	293	1	30	31	135
61	HELDY74	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	71	932	1	932	201	201	294	1	17	18	33
62	HEMAE80	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	72	996	1	945	12	12	295	1	24	25	136
63	HEEBA88	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	73	785	464	785	356	356	296	1	29	30	57
64	HFGAB89	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	74	1069	196	1047	295	295	297	1	32	33	34
65	HFVHY45	97975	pBluescript	75	831	1	831		89	298	1	30	31	76

Gene No.	cDNA Clone ID	ATCC Deposit Nr. and Date	Vector	NT SEQ ID NO. X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO. Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/04/97 209081 05/29/97												
66	HGBAJ93	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	76	590	1	590	233	233	299	1	38	39	94
67	HGBBQ69	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	77	1274	1	1273	105	105	300	1	24	25	43
68	HHFCF08	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	78	1133	4	1042	175	175	301	1	23	24	30
69	HHFHJ59	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	79	661	1	661	192	192	302	1	29	30	112
70	HHFHR32	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	80	1378	1	1378		358	303	1			13
71	HHGCN69	97975 04/04/97 209081 05/29/97	Lambda ZAP II	81	1440	298	1440	532	532	304	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig. Pep	Last AA of Sig. Pep	First AA of Secreted Portion	Last AA of ORF
72	HHGDO13	97975 04/04/97 209081 05/29/97	Lambda ZAP II	82	1381	766	1371	993	993	305	1	23	24	34
73	HHPEFD63	97975 04/04/97 209081 05/29/97	Uni-ZAP XR	83	1706	182	1644	257	257	306	1	24	25	81
74	HHSEG23	97976 04/04/97	Uni-ZAP XR	84	573	1	573	160	160	307	1	18	19	71
75	HJPV06	97976 04/04/97	Uni-ZAP XR	85	684	199	684	323	323	308	1	27	28	33
76	HKIXL73	97976 04/04/97	pBluescript	86	1036	591	1036	690	690	309	1	32	33	114
77	HKMNC43	97976 04/04/97	pBluescript	87	908	1	908	139	139	310	1	18	19	108
78	HMEJE31	97976 04/04/97	Lambda ZAP II	88	655	1	655	165	165	311	1	33	34	64
79	HMSKS35	97976 04/04/97	Uni-ZAP XR	89	1102	1	1102	228	228	312	1	26	27	49
80	HNF4E54	97976 04/04/97	Uni-ZAP XR	90	1533	665	1518	347	347	313	1	26	27	293
81	HNFJH45	97976 04/04/97	Uni-ZAP XR	91	575	1	575	275	275	314	1	30	31	67
82	HNGBT31	97976 04/04/97	Uni-ZAP XR	92	639	1	639	224	224	315	1	28	29	104

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO.	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO.	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HNGIN60	97976 04/04/97	Uni-ZAP XR	93	744	1	744	225	225	316	1	43	44	70
84	HNGJG84	97976 04/04/97	Uni-ZAP XR	94	526	1	526	268	268	317	1	29	30	38
85	HINHFW42	97976 04/04/97	Uni-ZAP XR	95	426	1	426	168	168	318	1	28	29	71
86	HNHFL57	97976 04/04/97	Uni-ZAP XR	96	844	1	844	98	98	319	1	25	26	61
87	HOGAR52	97977 04/04/97 209082	pCMVSPORT 2.0	97	1985	453	1985	533	533	320	1	17	18	285
88	HOSBZ55	97977 04/04/97 209082	Uni-ZAP XR	98	1416	69	1416	246	246	321	1	32	33	54
89	HOSDI92	97977 04/04/97 209082	Uni-ZAP XR	99	1935	141	772		274	322	1	20	21	58
90	HPBCU51	97977 04/04/97 209082	pBluescript SK-	100	599	1	599	86	86	323	1	27	28	119
91	HPCAL49	97977 04/04/97 209082	Uni-ZAP XR	101	784	1	784		280	324	1	18	19	43

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
92	HPFCR13	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	102	1035	602	1035	859	859	325	1	32	33	58
93	HPHAC83	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	103	2218	840	2182	1035	1035	326	1	17	18	17
94	HPMBQ32	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	104	1351	1	1351	18	18	327	1	23	24	86
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	105	2066	51	2052	270	270	328	1	29	30	537
95	HPWAN23	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	226	2057	1	1954	220	220	449	1	29	30	315
96	HRDFB85	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	106	1705	23	1697	233	233	329	1	21	22	201
97	HRGBR28	97977 04/04/97	Uni-ZAP XR	107	1167	611	1167	53	53	330	1	1	2	263

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		209082 05/29/97												
98	HSKGN81	97977 04/04/97 209082	pBluescript	108	1907	151	1432	353	353	331	1	23	24	260
		05/29/97												
98	HSKGN81	97977 04/04/97 209082	pBluescript	227	2084	335	2084	537	537	450	1	19	20	23
		05/29/97												
99	HSPAHS6	97977 04/04/97 209082	pSport1	109	611	1	576	229	229	332	1	25	26	47
		05/29/97												
100	HE8EU04	209746 04/07/98	Uni-ZAP XR	110	2632	294	2632	337	337	333	1	25	26	333
100	HSXBT86	97977 04/04/97 209082	Uni-ZAP XR	228	2143	53	1096	235	235	451	1			9
		05/29/97												
101	HSXCS62	97977 04/04/97 209082	Uni-ZAP XR	111	2249	1	1953	90	90	334	1	18	19	199
		05/29/97												
102	HTEFU09	97977 04/04/97 209082	Uni-ZAP XR	112	2198	228	2158	400	400	335	1			23

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
103	HTTEKM35	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	113	1043	40	1043	320	320	336	1	20	21	142
104	HTGEP89	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	114	703	1	703	285	285	337	1	29	30	94
105	HTGEW91	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	115	3684	526	1338	584	584	338	1	24	25	37
106	HTOEY16	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	116	1965	127	1915	202	202	339	1	27	28	38
107	HTPCN79	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	117	503	1	503		1	340	1	7	8	70
108	HTSGM54	97977 04/04/97 209082 05/29/97	pBluescript	118	1133	316	1069		423	341	1	12	13	84
109	HTSHE40	97977 04/04/97	pBluescript	119	1101	118	956	218	218	342	1	31	32	89

Gene No.	cDNA Clone ID	ATCC Deposit Nr. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
110	HTWAF58	97977 04/04/97 209082 05/29/97	Lambda ZAP II	120	282	1	282	137	137	343	1	25	26	48
111	HTWBY29	97977 04/04/97 209082 05/29/97	pSport1	121	2635	1593	2489	1654	1654	344	1	25	26	55
112	HUKFC71	209007 04/28/97 209083 05/29/97	Lambda ZAP II	122	994	1	932		272	345	1	15	16	221
113	HCE3Q10	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	123	1542	1	1542	143	143	346	1	25	26	63
114	HCEVR60	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	124	1390	82	1390	127	127	347	1	32	33	153
115	HDI1AW95	209007 04/28/97 209083 05/29/97	pcMVSPORT 2.0	125	1288	412	1288	571	571	348	1			16
116	HEGEL90	209007	Uni-ZAP XR	126	1517	1	1452	243	243	349	1			9

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/28/97 209083 05/29/97												
117	HELB029	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	127	1073	198	1073		776	350	1			13
118	HERAH36	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	128	300	155	300	202	202	351	1			17
119	HFXBW82	209007 04/28/97 209083 05/29/97	Lambda ZAP II	129	1275	1	1275	56	56	352	1	23	24	61
120	HHP1D20	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	130	472	51	472		243	353	1			32
121	HIBED17	209007 04/28/97 209083 05/29/97	Other	131	1950	284	1927	395	395	354	1	72	73	245
122	HLTER03	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	132	990	1	990	78	78	355	1	22	23	34

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
123	HOABL56	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	133	1720	565	1720	660	660	356	1	18	19	21
124	HPMCJ92	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	134	705	28	705	106	106	357	1	28	29	98
125	HPWAZ95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	135	323	1	323	88	88	358	1	27	28	78
126	HRGBR18	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	136	582	1	582		16	359	1	17	18	30
127	HSUBW09	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	137	1021	1	1021	153	153	360	1	32	33	56
128	HUKCOG4	209007 04/28/97 209083 05/29/97	Lambda ZAP II	138	1777	439	1777		521	361	1			2
129	116EAA53	209007 04/28/97 209083	Uni-ZAP XR	139	643	303	643		313	362	1	7	8	31

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
130	HAGAI11	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	140	1220	1	1220		127	363	1	16	17	27
131	HAGAO39	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	141	721	1	721		415	364	1			14
132	HALSK07	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	142	1468	125	1468	210	210	365	1	29	30	33
133	HALSQ59	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	143	300	4	300	101	101	366	1	22	23	66
134	HAIBP89	unknown 05/18/98	Uni-ZAP XR	144	2243	173	2243	311	311	367	1	27	28	317
134	HBCGB91	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	229	1025	409	1025	624	624	452	1	20	21	25
135	HBMID81	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	145	1082	163	1082	357	357	368	1			30

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
136	HBXGK12	209008 04/28/97 209084 05/29/97	ZAP Express	146	4313	1153	4313	1313	1313	369	1	18	19	42
137	HFKEJ07	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	147	1183	1	1183	149	149	370	1	41	42	254
138	HCOA140	209008 04/28/97 209084 05/29/97	Lambda ZAP II	148	734	1	734	285	285	371	1			19
139	HCWHZ24	209008 04/28/97 209084 05/29/97	ZAP Express	149	1405	1	1405	108	108	372	1	34	35	63
140	HE2GT20	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	150	2890	1178	2890	1178	1178	373	1	31	32	39
141	HE8EY43	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	151	2399	1181	2399	1265	1265	374	1	30	31	34
142	HFCB37	209008 04/28/97 209084	Uni-ZAP XR	152	802	352	802		487	375	1			10

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
143	HFTCT67	209008 04/28/97	Uni-ZAP XR	153	461	24	461	145	145	376	1	37	38	63
		209084 05/29/97												
144	HGLAM46	209008 04/28/97	Uni-ZAP XR	154	2388	818	2388	648	648	377	1			18
		209084 05/29/97												
145	HHGBR15	209008 04/28/97	Lambda ZAP II	155	642	322	642	400	400	378	1			4
		209084 05/29/97												
146	HJAAU36	209008 04/28/97	pBluescript SK-	156	1251	583	1251		933	379	1	16	17	16
		209084 05/29/97												
147	HUSTF49	209008 04/28/97	pSport1	157	2127	247	2127	383	383	380	1	47	48	83
		209084 05/29/97												
148	HKLAB16	209008 04/28/97	Lambda ZAP II	158	1625	817	1625	1012	1012	381	1	18	19	20
		209084 05/29/97												
149	HLMMU76	209008 04/28/97	Lambda ZAP II	159	1687	1307	1687	1296	1296	382	1	28	29	28

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
150	HMSKQ35	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	160	1842	172	1463	319	319	383	1	30	31	33
151	HNHED86	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	161	770	1	770	30	30	384	1	31	32	46
152	HNHE188	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	162	519	1	519	242	242	385	1	17	18	24
153	HNHFQ63	209008 04/28/97 209084 05/29/97	Uni-ZAP XR	163	753	1	753	164	164	386	1	17	18	67
154	HOECU83	209009 04/28/97	Uni-ZAP XR	164	1400	189	1400		508	387	1	22	23	33
155	HPTRC15	209009 04/28/97	pBluescript	165	2153	594	2153		611	388	1			13
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	166	1251	219	1120			389	1			
156	HSKCP69	209009 04/28/97	Uni-ZAP XR	230	1250	223	1250	393	393	453	1	32	33	171
157	H6EAE26	209009	Uni-ZAP XR	167	882	48	882	155	155	390	1	33	34	153

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
158	HAGBX03	209009 04/28/97	Uni-ZAP XR	168	1208	1	1208	182	182	391	1			8
159	HAGDQ47	209009 04/28/97	Uni-ZAP XR	169	1307	1	1307	44	44	392	1	22	23	60
160	HAICP19	209009 04/28/97	Uni-ZAP XR	170	1624	89	1483	128	128	393	1	18	19	446
161	HAUAE83	209009 04/28/97	Uni-ZAP XR	171	2003	889	2003	1080	1080	394	1			23
162	HBHAD12	209009 04/28/97	Uni-ZAP XR	172	786	1	786		176	395	1	17	18	23
163	HBMIV28	209009 04/28/97	Uni-ZAP XR	173	1758	962	1758	1184	1184	396	1	27	28	34
164	HBMVPO4	209009 04/28/97	Uni-ZAP XR	174	888	330	862		546	397	1			2
165	HCDDDB78	209009 04/28/97	Uni-ZAP XR	175	2379	750	2379	901	901	398	1	18	19	24
166	HCEQA68	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	176	1348	1	1348	12	12	399	1	28	29	78
167	HCEZS40	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	177	1502	178	1502	315	315	400	1			20
168	HCFNF11	209010	pSport1	178	1637	26	1607	152	152	401	1	44	45	257

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/28/97 209085 05/29/97												
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	179	2911	1103	2858	192	192	402	1	32	33	424
169	HCRBL20	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	231	1811	20	1811	93	93	454	1	36	37	95
170	HCUBL62	209010 04/28/97 209085 05/29/97	ZAP Express	180	519	1	519	57	57	403	1	28	29	32
171	HDSAP81	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	181	968	320	968	476	476	404	1	27	28	79
172	HE2CT29	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	182	1128	1	1128	111	111	405	1	26	27	94
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	183	2276	48	2276	88	88	406	1	37	38	257

Gene No.	cDNA Clone ID	ATCC Deposit No. and Date	Vector	NT SEQ ID NO.	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO.	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
173	HE8MG65	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	232	2271	56	2232	79	79	455	1	43	44	170
174	HE9FB42	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	184	2500	76	1693	518	518	407	1	1	2	623
175	HEMAM41	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	185	1337	60	1328	175	175	408	1	39	40	190
175	HEMAM41	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	233	1338	33	1327	175	175	456	1	32	33	91
176	HEMCV19	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	186	941	33	931	79	79	409	1	23	24	178
177	HEMDX17	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	187	654	1	654	137	137	410	1			13
178	HEIAR54	209010 04/28/97 209085	Uni-ZAP XR	188	1848	454	1848	948	948	411	1	14	15	232

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
179	HETBX14	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	189	1146	157	1146		74	412	1	14	15	53
180	HFGAB48	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	190	906	156	906	245	245	413	1	30	31	32
181	HFKF140	209010 04/28/97 209085 05/29/97	Uni-ZAP XR	191	1941	120	1002	213	213	414	1	18	19	218
182	HFXHN68	209010 04/28/97 209085 05/29/97	Lambda ZAP II	192	2118	777	2118	966	966	415	1	23	24	50
183	HGBFO79	209011 04/28/97	Uni-ZAP XR	193	1538	259	1538	273	273	416	1	23	24	49
184	HGLAM56	209011 04/28/97	Uni-ZAP XR	194	1098	68	1098		185	417	1	28	29	69
185	HHLBA89	209011 04/28/97	pBluescript SK-	195	1001	1	1001	324	324	418	1	25	26	39
186	HHPDW05	209011 04/28/97	Uni-ZAP XR	196	1443	1	1443	246	246	419	1	21	22	21
187	HHPDSD37	209011 04/28/97	pBluescript	197	1282	66	1282	171	171	420	1	19	20	37

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO:	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO:	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
188	HHPSF70	209011 04/28/97	pBluescript	198	951	26	951		162	421	1	16	17	34
189	HHSAK25	209011 04/28/97	Uni-ZAP XR	199	1740	1390	1740	1534	1534	422	1	19	20	31
190	HIASB53	209011 04/28/97	pBluescript	200	1707	401	1195	652	652	423	1	26	27	126
191	HIABZ65	209011 04/28/97	pBluescript SK-	201	779	1	779	23	23	424	1	26	27	68
192	HIPIBB39	209011 04/28/97	Uni-ZAP XR	202	1617	188	1605	182	182	425	1	28	29	91
193	HLHSK94	209011 04/28/97	pBluescript	203	1974	1	1794	112	112	426	1	26	27	379
194	HLHTC70	209011 04/28/97	pBluescript	204	1057	229	1057	365	365	427	1	23	24	22
195	HLMIW92	209011 04/28/97	Lambda ZAP II	205	721	1	721	244	244	428	1	25	26	46
196	HLTCY93	209011 04/28/97	Uni-ZAP XR	206	2465	988	2465	1225	1225	429	1			4
197	HLTIDB65	209011 04/28/97	Uni-ZAP XR	207	1480	1	1480		371	430	1	15	16	143
198	HMSHM43	209011 04/28/97	Uni-ZAP XR	208	872	1	872	35	35	431	1	18	19	36
199	HMSHQ24	209011 04/28/97	Uni-ZAP XR	209	1779	16	1779	148	148	432	1	24	25	36
200	HNF-AH08	209011 04/28/97	Uni-ZAP XR	210	2110	592	2110	611	611	433	1	18	19	191

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
201	HNGAO10	209011 04/28/97	Uni-ZAP XR	211	938	1	938	107	107	434	1	27	28	30
202	HNGBE45	209011 04/28/97	Uni-ZAP XR	212	1551	1	1551	114	114	435	1	21	22	100
203	HNHAZ16	209011 04/28/97	Uni-ZAP XR	213	997	1	997	202	202	436	1	24	25	36
204	HNHCM59	209011 04/28/97	Uni-ZAP XR	214	1496	1	1132		165	437	1	28	29	41
205	HOSFM22	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	215	1308	501	1308		809	438	1			1
206	HPHAC88	97977 04/04/97 209082 05/29/97	Uni-ZAP XR	216	1705	384	1705	549	549	439	1	23	24	24
207	HCDEO95	209007 04/28/97 209083 05/29/97	Uni-ZAP XR	217	999	608	999	273	273	440	1	22	23	54

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table I. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

5 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources
10 using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the
15 cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1
20 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra.*) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide
25 was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results
30 shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., +
35 or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

10 Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to 20 a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between 30 a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a 35 FASTDB alignment of DNA sequences to calculate percent identity are:
Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30. Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,
5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be
10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and
15 subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window
20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.
25 For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of
30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are
35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

5 The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

10 The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

20 The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

25 The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

30 Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4° C or frozen at -80° C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then
5 be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA
10 synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic PI library (Genomic Systems, Inc.) is screened by PCR
20 using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by,
25 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is
30 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are
35 mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μ l of reaction mixture with 0.5 μ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the fl origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the fl ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lacmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ³²P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).)

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafmid BA	plafmid BA
pSport1	pSport1
pCMVSPORT 2.0	pCMVSPORT 2.0
pCMVSPORT 3.0	pCMVSPORT 3.0
pCR [®] 2.1	pCR [®] 2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Altting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Altting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

5 Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined
10 from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95%
15 identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

20 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

25 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous
30 nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under
5 stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which
10 comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide
15 sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at
20 least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

25 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide
30 sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer
35 as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X, wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

5 All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

10 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with
15 a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

20 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic
25 surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian
30 rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

35 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase
5 regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue
10 regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and
15 peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease,
20 Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

25 A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular
30 trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body.
35 For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria.

5 Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

10 Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

15 These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or
20 diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo
25 therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to
30 differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion
35 injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

5 Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Bimaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, 10 Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., 15 Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, 20 Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that 25 can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, 30 Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, 35 and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases:

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human
5 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The
10 Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene
15 expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to
20 supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired
25 response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such
30 as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a
35 recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (^{125}I , ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^3H), indium (^{112}In), and technetium ($^{99\text{m}}\text{Tc}$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, ^{131}I , ^{112}In , $^{99\text{m}}\text{Tc}$), a radio-opaque substance, or a material detectable by nuclear magnetic

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein

Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available.

As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.

Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

5 Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

10 Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final
15 preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

20 Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG)
25 can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

30 Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the
35 fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the

carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a
5 stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion
10 (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280}
15 monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from
Commassie blue stained 16% SDS-PAGE gel when 5 μ g of purified protein is loaded.
20 The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

Expression System

25 In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient
30 polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that
35 express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase: *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μ g of a plasmid containing the polynucleotide is co-transfected with 1.0 μ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One μ g of BaculoGold™ virus DNA and 5 μ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm

tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

5 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)
10 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in
15 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection
20 ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ³⁵S-methionine and 5 µCi ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins
25 in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

30 Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates

the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV I, HIV I and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the

polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five μ g of the expression plasmid pC6 is cotransfected with 0.5 μ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μ M, 2 μ M, 5 μ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 μ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

```
GGGATCCGGAGCCCAAATCTTCTGACAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
CCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
```

AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
5 GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCACCGTGGACAAGAGCA
GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
10 GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of
15 the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

20 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell
25 Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at
30 about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line
35 (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as

described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20:

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a

working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The
5 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2×10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (I2-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x
10 Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression
15 vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a
20 control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of
25 cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl_2 (anhyd); 0.00130 mg/L
30 $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; 0.050 mg/L of $\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$; 0.417 mg/L of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 311.80 mg/L of KCl; 28.64 mg/L of MgCl_2 ; 48.84 mg/L of MgSO_4 ; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO_3 ; 62.50 mg/L of $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$; 71.02 mg/L of Na_2HPO_4 ; .4320 mg/L of $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic
35 Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of

Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0
5 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H₂O; 99.65 mg/ml of L-
10 Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine;
15 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mM glutamine and 1x
20 penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B
25 adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

30 It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an
35 activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u> <u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>	<u>STATS</u>	<u>GAS(elements) or ISRE</u>
<u>IFN family</u>							
5	IFN- α /B	+	+	-	-	1,2,3	ISRE
	IFN- γ		+	+	-	1	GAS (IRF1>Lys6>IFP)
	IL-10	+	?	?	-	1,3	
<u>gp130 family</u>							
10	IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	IL-11(Pleiotrohic)	?	+	?	?	1,3	
	OnM(Pleiotrohic)	?	+	+	?	1,3	
	LIF(Pleiotrohic)	?	+	+	?	1,3	
	CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
15	G-CSF(Pleiotrohic)	?	+	?	?	1,3	
	IL-12(Pleiotrohic)	+	-	+	+	1,3	
<u>g-C family</u>							
	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
20	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
	IL-15	?	+	?	+	5	GAS
25	<u>gp140 family</u>						
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
	GM-CSF (myeloid)	-	-	+	-	5	GAS
30	<u>Growth hormone family</u>						
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
35	<u>Receptor Tyrosine Kinases</u>						
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCCGAAATCTAGATTTCCCGAAATGATTTCCCG
AAATGATTTCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4).

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCCCGAAATCTAGATTTCCCGAAATGATTTCCCGAAATG
ATTTCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGC
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final
5 concentration of 10^7 cells/ml. Then add 1 ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

10 On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

15 Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12
20 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples
25 from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

- 10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^7 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

- 15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM MgCl_2 , and 675 uM CaCl_2 . Incubate at 37°C for 45 min.

- 20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

- 25 These cells are tested by harvesting 1×10^8 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

- 30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two ml of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 µl per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 µg/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 µg/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating

diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCCA
TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT
AATTTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC
CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 μ l of 12 μ g/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 μ l of buffer.

5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 μ l of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100
10 μ l/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 μ l, followed by an aspiration step to 100 μ l final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the
15 following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 μ l. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

20 Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase
25 RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members
35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇ and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

5 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
10 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction
15 mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-
20 POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of
25 tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine
30 kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,
35 Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenin deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. USA* 82:3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily
10 dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5,
20 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is
30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier: P.T. et al., DNA. 7:219-25 (1988)), flanked by the long
5 terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set
10 forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to
15 transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is
20 then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media,
25 containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is
30 required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph. blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for the polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with

liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

Sequence Listing

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Human Genome Sciences, Inc., et al.
- (ii) TITLE OF INVENTION: 207 Human Secreted Proteins
- 10 (iii) NUMBER OF SEQUENCES: 800
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue.
- 20 (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: USA
- 25 (F) ZIP: 20850
- (v) COMPUTER READABLE FORM:
- 30 (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
- (B) COMPUTER: HP Vectra 486/33
- 35 (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text
- 40 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 45 (B) FILING DATE:
- (C) CLASSIFICATION:
- 50 (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER:
- 55 (B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Kenley K. Hoover

(B) REGISTRATION NUMBER: 40,302

(C) REFERENCE/DOCKET NUMBER: P2007PCT

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 733 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

55	GGGATCCGGA GCCAAATCT TCTGACAAAA CTCACACATG CCCACCGTGC CCAGCACCTG	60
	AATTCGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCAAA ACCCAAGGAC ACCCTCATGA	120
35	TCTCCCGGAC TCTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG	180
	TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG	240
40	AGGAGCAGTA CAACAGCAGG TACCGTGTGG TCAGCGTCTT CACCGTCTTG CACCAGGACT	300
	GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG	360
	AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCGGAGAACC ACAGGTGTAC ACCCTGCCCC	420
45	CATCCCGGGA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT	480
	ATCCAAGCGA CATCGCCCTG GAGTGGGAGA GCAATGGGCA GCGGAGAAC AACTACAAGA	540
50	CCACGCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG	600
	ACAAGAGCAG GTGGCAGCAG GGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC	660
	ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCGGG TAAATGAGTG CGACGGCCGC	720
55	GACTCTAGAG GAT	733

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
1 5

10

15 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

25 GCGCCTCGAG ATTTCGCCGA AATCTAGATT TCCCCGAAAT GATTCCCCG AAATGATTTC 60
CCCGAAATAT CTGCCATCTC AATTAG 86

30

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

40 GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

45

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

55 CTCGAGATTT CCCCAGAAATC TAGATTTCCC CGAAATGATT TCCCCGAAAT GATTCCCCG 60
AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAAGTC CGCCCATCCC 120

60

268

GGCCCTAACT CCGCCAGTT CCGCCATTC TCGCCCCAT GGCTGACTAA TTTTITTTAT 130
TTATGCAGAG GCGGAGGCG CCGCGCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240
5 TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10 (2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

20 GCGCTCGAGG GATGACACCG ATACAACCCC GG 32

25 (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

35 GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

40 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

50 GCGGACTTTC CC 12

55 (2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid

269

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

5 GCGGGCTCGA GGGGACTTTC CCGGGGACTT TCGGGGACT TTCCGGGACT TTCCATCCTG 60
CCATCTCAAT TAG 73

10

(2) INFORMATION FOR SEQ ID NO: 10:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTTCCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60
25 CAATTAGTCA GCAACCATAG TCCCGCCCTT AACTCCGCCC ATCCCGCCCC TAACTCCGCC 120
CAGTTCGGCC CATTCTCCGC CCCATGGCTG ACTAATTTT TTTATTTATG CAGAGGCCGA 180
GGCCGCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTGTG GAGGCTAGG 240
30 CTTTTCGAAA AAGCTT 256

35

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2526 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

45 GACAGGCTAT CCGAGAATCT GAGAGTGGG CCGGCAATT CCTCCAGTGA CCCTTGTGAC 60
CTAAGTCCAG TCACACATTT CCCAAAGTTT CTCCTTGTCA TAACCCCTGGT CTGGCTGGTT 120
50 TTGRRGRTT GAGAATGGGT CAGGGACTCC AGGCCAAGTC CAACAGAGAC CCCAAACCCA 180
CCACACACCA GCAGCCACAA CCTCACCACC AACAAAGAGG ACTTTTGTGG GGCCACAAGT 240
AAGAGTTCAT TTCTGGAATG GACTCAGACC TTAAACAGG AGAGTTGAGC ACTTCCAGKS 300
55 AGTTTPTAAG CAAGGCATGG GGAACAGGA ATAGAACCTT TCAAAGAGGT TGCCGAGAGA 360
AAAGCTGGGC CTCTTGCAAT CGGCTTCCTT GGAGCAGCCT CTCTGGCAG AAAGCCATCA 420
60 GGTGCTCAAT CATCTCTCC TGGCCAAAGG TGTGACCATG CTTAGTACTG GAATAGAGGT 480

	GGCCAGGGCC CCAGCGACTC TTCTTGGCCT GATGTTTGTG CTCACAGGCA TGCCACGTGG	540
5	CCTGAGATGA TTCAGAACAA ATCATGCTAA CTTTGAATCC ATCCAGCCAC TTGCAAATGA	600
	TAATCAGAAG TCAGCTTGTT CACTGTTAGA AAGAACTAA CAAAAGAGAA CCCAGAGCAA	660
	TCTAGAATCT TTGAGTGCTT GGCTTTCCAA GGATACTGCG GAGACTCTGG CCAAGCTGAT	720
10	GAMCTTCTGA ARTGTCACTG GCACCATATG CAACAAGAAC CACCATTAC TGAGTAGCTA	780
	ATGGGTTTGG GGCTGGGAC ATTCCATCTG AGGTCTTCC TGAACATGTC ACTCCACAGC	840
15	AGAGGACCCG TTGCAGCTTA CCCAGAACCA CTCCTCCAGG AGAGCTGGAT GTTTTGCCTG	900
	CAACACCTTG AGCACTGACT GCTATTGTTT AAAAAAGCC TTGCTGCAT TCGGAGGACT	960
	GGCCCGTGCC CTGAGGTGAC TTCCTAACTA TGTGGTTTCA TTAGCGAATT TATTTTGT	1020
20	GCTGGGTGGA CATTTGTATT TTGTTAGGTT GCTGTTTAAG CTCAGTTTG CTGTGCTCTC	1080
	TGCAGCTACA AAACATCTTG GCATATTTAA GAKTGGCTTT <u>TATAAATAGC</u> TTTATTCTGA	1140
25	TATTAATCAG ATTCCCACT TTAGTGAGAA TTAAGGACTG GGGTACTTTA AAGAAATGCA	1200
	AATAGCAATT GAAGAACCAC TGCTGCAGGT GGTAGCCCTG GCTAGACTGA ATTACACTAG	1260
	AAATCAGCCA GAAGGAAGCG TCCTTGGGAT CCCAGATCAC TCTTTTTTTT TTTTTTTTA	1320
30	AAAGGGGCG CCCCTTGATG GCTCATCTCT CTGAATAACA GTTACGTCTT CATATCGATA	1380
	CCAGATGCCT TCTTCATCAT GCCACTGAAG CCACTCACCA CTTTCAAGAA CATGCCAACC	1440
35	TCTGTCAGAT TCACTTACCC ACABACAAGG AGGCACGTTT GGCACAAAGT GTTGTCTCTC	1500
	AGGTCCAAGT GGACTCTACA GAGTGCTTGA CCTCAACACA CTGGATTCCA GGTGGACTGG	1560
	ACCAAGAGCA GCGAAAGACA CGGGAAGTGA AAACTCCAC AGGGTTTGGG GAATAGAAAT	1620
40	GAAAGCCAC GTCATATAAC TCAAGAATAA ATGGTGTGTT GGAAATTTTA AAATTATCAT	1680
	CGAAGGTGGT GAACTATTT CAGGCCCAA TGAAAGGAAA TCGCCAGTTG GGGATGAAAT	1740
45	CACAGAGCCT GTGTTTTATG ATATGGTTGG ATGTCCACTG ATGAAATTTT AAAGGAGTTT	1800
	CATTTTAAA AGTGGCATG ATTCTACATA TGAGAATTCT TTAGGCCAAG AAAGTGTCT	1860
	TGGCTCAGAG GTGTTGGGAA TTAAGCAGA GAGAAGCCAT TCGTGATGCT TAGAACCAG	1920
50	GATGGTCATG TACACAAAGA CCATCGAGAC GGCCATTCTT GTTTACAAAA CACTTACCAA	1980
	GAAAGCACTT TGTAGGGGAA CTTTAGTAAG TTCTTCTCAT TTCATTATGT TTCTTCCAAG	2040
55	GAAACAGGAG AGACTGAATT AATAATTCTC TCTTTCTCTT TAAGCACTTT TAAAATAATA	2100
	AAGTACATCT TGAAATTTGG GGGGCGATCT CTGATTTAAA AAAAGAAAAA GGCTGCTTGA	2160
	TGTATGTTAT GCAGAGACAC TCTGCTCTG GTGGCTGCAG AGCAATACCC AAGCCTCAT	2220
60	TGGAAGGCTC AACATTGGA ATTGCACTTT AATTGATTAA TCCTCAATTC ATGTGCTCTT	2280

ACGGGATGGT GGGTCTGGGA CCCCAATCA TTCTTATCTG CCAAAGAATT ATCTAGAAGC 2340
ACATCAATA CCAGCACCCC ACCTGCACAA TGGGGGTGGA AAACCTTTTGT ATCCCTAAGC 2400
ATATTATTTT ATAGTGTCTG CCATGCCATG TGGAAATACT TTATTTTAA CCTCAGGATT 2460
TAAATAAAGT AAACACTATG ACATTTAAAA AAAAAAAAAA AAAACTCGAG GGGGGCCCCG 2520
TACCCA 2526

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1131 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CACTGCACCA GCTTGTAT CTGTAAAATG ATGATAATAC CAACACCTTC TTCTTGGGGT 60
ACTGAAGATG AGAGAACATG ATATGTGTAA AGTGCCCTTC ACAATACCCA GAACATAGCA 120
AACATGTAAT GAATGTAGTA ATAGTAATTA TTTTATTTTC TTTTGATTCA GTTGGGACTA 180
TGTTCAGCTG TAACAGAATA CCCAAAATAA CTGTTTTAAA CAAATTAAG TTTWGTGTG 240
AAGTTTTGTT ACGAATTCAG ACAATCCAGG GCTTTTATAG ATGCACCAGG ATCAGCAGGT 300
ACAAAGGCAT CTTTCTGAT TTCTGCCAGT CTCAATGCAT GGGTTGCAAT CCAGARTCCA 360
RGATGGCAGT TCCAGCCCTG GTTACGCCCA TATTAGCACA CAGAAAGAAA GAGAAAGGGA 420
TGTGCTCTT CACTTTAATC ATAGCTCCCA CTAGATGCAC CCACTACTTC TGCTGATACT 480
CCATTAGCTA ATGCTTGCTT ACATGGTCAC ACTTAGTTTC CAGAGAGACA TGTCTGGACA 540
GTCATGTGCT CAATTAAAT CCAAGTGTCC AATTACTGAG AAAAAAGAA ACTAGCACCT 600
TTGCTTGGTT GCATTCTCT TAGCATAAGC CACATTCTTT TTATGAAGTT GTCCTCAGTT 660
ACTTGGATGC CTCAGTTGTC CTTTCAWTTA GAAAWGCYCC TKGACAYCC TGAAWCTGAC 720
TTCTTTTGTG ATCAGCACCA TCACTACCAC TGGCTTCTTC AAAGCCACCA CTTTCTGTCC 780
CCAGGATGGT TGCAACAACC ACCATAGGGA CTTTTTGCCT TCTACTTCCA CAGAAAGNC 840
CAGAGTAAGC TTTTGAAAAT GTAGGTGAGA TCATGTCTCT CTCTTCTCT TCAAAACCT 900
CCCGATGGCT TTTCAATTA CTCAAAAGAA AACCTAAAAC TTTGCTGTGA GATCTATGTG 960
ACCGGCTTA TTCTTCTCT TACTTTATCT CTGTATGCT CTTCTCACT CTACTCCAGC 1020
CATCCACCT CTTTGTGCT TGTCTTATAC TCCTAAAAGA AGTTCAGTCT TCCCTTATGA 1080

272

TATTTGCACT TAAATAGAA AAAAAAAAAA AAAAAAACT CGAGGUGGGC C

1131

5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 941 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

20

GGCAGGAGTA GCATTTCAAT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTCATGT 60

GATGTCCTCA GCCAAGATTG TRAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA 120

GGCTGGAGAG ATCATATTTT TGGTATTAAA CTGGAGTCTC TCCATCCTTC ACATTGTTGA 130

TGTCCTCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCGGCTAGCG GTTTGAGCCA 240

GAGAATGACA GCTCTGGTTT GGAGAAAAGG GCGGATGGT GGCTCTAGAA AGCCCATCCT 300

TCTGCTCTTC TTTTCTCTCC CCTTATATT GTGCTTTCAI TCATTCATTG ATTCATCAAA 360

CATTTGTTGA GCACCTATTA TGTGTCAAGC TGTGTGCTAG CCTCTGAAA ACCTGCCCTC 420

ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CGGTTGTCA 480

GGGTCTCACA GAGCAGTGGC CCCTCATCCA GACCGATGAG GTCAAAGAAG GCATCCAGGC 540

GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGAGGGAGCT GCACCASCAG GGGTTGGAAC 600

TGAAGGTGGC AGTGCTTGA GTCTTGATTG CAGCAGAGGG AGAGCAGTCT GTGAAAAGGC 660

ACCAAGGGTG GGAGAGGGCA GAGCACATGG AGGAACCTCA GGTAGTTCTG GATGGCCTG 720

GGGCAAGCT AGAGAGCTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAACCTCCA 780

TCCCAATAAA CCCATTGGAA ACGAAAAATT TAAGTCAGAA GTGCATTTAA GGCTGCTCCG 840

AGTAGAATGA TTTTACAAC GAATTGATCA CAACCAGTTA CAGATGCTTT TGTTCCTTCT 900

CCACTCCCAC TGCTTCACCT GACTAGCCTT TAAAAAATA A 941

50

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

55

(A) LENGTH: 843 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

	CTNAGGGATAA CCCCAGAGT GGGAAATAAA CCTCAATTA AAGGGGGAAC CAAAAGCTG	60
	GGAGTTTCCC CCCCCGGTG GGGCCNGNT CTAGGAATA GTGGAAATCCC CCGGGGCTGC	120
5	ACGGAATTCG GCACGGAGTG GGAATCTTGT TTGTATGATA CTATTTCCAC AAWATGCATT	180
	GAGACTTGGT KTGTGGCCTA GGACATGGTC AATCTTTTAT AAATATTCGG TGAATTTCTT	240
10	TAGTGCATAT TCTCCGATGG GGGCTGTGGG GACAGAGTTC TAAATATGCC CATTAGATTA	300
	AATCTCTTCA TTCTGTTGCT CACATCTTCT ATATCCTTAT TAATCTGTCA ATCTCTTCAA	360
	GAGAGGTGTT ATTAAAATCT CTCACTGTAT GTGTCACTTT GCCCTTAAAA TTCTGATGAT	420
15	TTGCTTTATA AATGGTTATA ACCATTTTCC AGGAAGAACA TTAAAGAACT TTCCATTGGC	480
	ATTATCCAGT TTCCCTCAAA ATACTGGTTT TTTTATTTT GGCTNCTAAG CAGCTATGAA	540
20	TCGAGTTTCT CAGAAGCCCT TGTCTCAAGG CATTTGTTTC CAGATTACCT TGTTAGCATC	600
	CACACTATGG GCTATTTTAG AAAACAAAA AAAGTATCAA AATCATATAG CTATGATTTT	660
	CCTGTGCTTG AAGGAGCCTT AAAGCTCATC TAGTCCAGCC AGTATTTGTT CATCCAAAT	720
25	CTGCCAAGAA ATCTCTATTG TCAAGATATT CTMTACCATC TTTGGGACAT TCTCATTATT	780
	AGAAACAAAT CCTAAGAAGA AATTCTGCCA TAKACAACCC ATCCGTCTTT TAAAAAATA	840
30	AAA	843

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1018 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45	CTGTAATTTT TAATTTTCAT ATACCGTGCT TTGATCTTAA TTTTATTTT TGAGTTCTCT	60
	GAAGGTTACA TATACAGAGT GCTTCAGGAA TGATCATTTT GTTATTATTC ATGCTTCTTA	120
	ACAATGTTGT TTTAGTCCAA GAAGATAATT GCCAGAGAAA GAATACAGTC CAGGAAAGAA	180
50	GARGCTGGAG CCAGTGGTGA AGARGGATTG AGARGACAGA CATTGTGGGA ATGAAATCAT	240
	GAATAATCGT GTTTTGAAT TGTCCAAAAA CTCTACAAA CCATGAAATG TTGGAGTTTA	300
55	AATCTAATTG TTGAAAAAT CCCCACATTC CTGTATGCC TTAGGTTGAG CATAATTCCA	360
	CATCCGTGGA CTGATGCACT TCCCAAGAGG GGGCTCATTT AACTCTTCCG AGGCAGCAGC	420
	AGCAAGGGCA CCCCCTCCTT TCCCCCAACA CCCCAATTCT CATGGCTCTT CTTTCTCTCA	480
60	TCTCATGCTT AGGTTAGAAA AGGGCACAAG GTAAGGAAGC CCTTGGGAAT AGGCTGAATC	540

TGGCTATCTA ATTTGGTGGC AAATACTTAA TGTGCTTGAA TTTAAAAACA GCRAACATGT 600
 AGAAAGGTAA TTATAATTAT GAGGCCAGTT CTTTAAGCTA GCTTTTTTTC CCCTCTCAAA 660
 5 CAGCATATTG GCTTGGATGT CAGCAGGAGA AAGTGTTTTT TGCAATACAC ATAATGCATA 720
 TATGGTCCTG TTAGCAATCT ATAGAAAATA GATATTGCTC ATTAAGGTAA ATATTTTGT 780
 10 TGATGAATGA TCTGGAATGG TCTGGACTTG TTGTGTGAAC AGGAAATTGC TCTGTAGGCT 840
 TTGACTTGTG AGCTAAAGAG TGAGGCTGGT AAGATTAATT AAAGTAAATA CTGTGACAAT 900
 AGGATGTCAA AACCAAAAAC GTGTTTGTGA AACTCAAGGA ATTAATGACA CATAGGGAAG 960
 15 TTTTGGCTAT ATTAAGCATA GAGTAGGAGA GCCAAGTCAA GAATAAAAAA AAAAAAAA 1013

20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 661 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

30

TTTAAGAAAT TAGTGAATCC CCGGNTGCAG GGAATTCGGC ACGAGGAGGA GGGCGTCAGC 60
 TGGCAGGAGC GCAGGATGGC AGCTGYTCCC CCGGTTTGCA CCCCCCAGT TCTGCTGGAC 120
 35 ATAAGYTGCT TAACAGAGAG CCTGGGAGCT GGGCAGCCTG TACCTGTGGA GTGCCGGCAC 180
 CGCCTGGAGG TGGCTGGGCC AAGQAAGGGG CCTCTGAGCC CAGCATGGAT GCCTGCCTAT 240
 GCTGCCAGC GGCCTACGGC COTCACACAC CACAACACTG GCCTMTCCGA GCTGCTGGAG 300
 40 CATGGAGTGT GTGAGGAGGT GGAGAGAGTT CGGCGCTCAG AGAGGTACCA GACCATGAAG 360
 GTGGCCAGGG CAGGGCTCGG ACCTACCCCA GGAATGTCCT GCCCTGGGAA TGACAACACA 420
 45 GTCCACACCA TGCACGGGGA GGCAACAGG GGCAGCTGAC CCAGCCCAGG GGTCAGANGA 480
 GGTCTTGCCG AGGAAGTGGC AGCTAAGCTG ATACCTGATA TGCACWAGKC AGCCARGYGG 540
 AGACAGGCAA GGAAGAAGCT TGTTTTGAGG ACAGAATTTT CTAGATCACT CAGCACCATC 600
 50 TGGCTTTTGG GGCTTTTGT TTTATTTTGT TTTTGAGACG GGTCTCGCT CTGTGCCCCA 660
 N 661

55

(2) INFORMATION FOR SEQ ID NO: 17:

60

(i) SEQUENCE CHARACTERISTICS:

275

- (A) LENGTH: 553 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGCACAGGGC TATTTGCCCC TCTCTCCACA TGACAGAAGT GCTCTAAGTT TCTTTGCTGC 60
 10 TCTTCTCAGC TGTGAGACGG CTGCTGCTTT GTTTTCCACA CCACCATGTC TATTCTTTGC 120
 TGTCTTTWAC TCTGCTGTT TTTTTCCTTT TGTATTCTTT CTGGCTCTTG TCCCTTTTCC 180
 15 CACGTGTCWC AGCTTTCCTT TATTGCCACT TTCAGTCAGA GCAGTCTGT GCTTCTGGTG 240
 CCGGCATACA ATAATTACTT GAGTTTCTTG GCTTTTCTTG ACTGTGCTC TCTTACTTCA 300
 ACATAGGAAT AGCCTGTCTT AGAATTCTC CAGTTCCAGG GCTCAAGAGG GAGAGTGCCA 360
 20 GAAATTTGAG ACTGTTTTCC CTGTCTTGA TTGAATTCAT AAAGCAAAC CAGTGTTTGT 420
 GTGAGGGTTT GCTGTCTCAT GCCTATAGCT TGTTCGGTG CAAACCTATA GAATCCAGCC 480
 25 TCGGAAAAGA AAGRAACCAG AGAATANCAG CATCAGAACA ATGCTTGACA TCATTTCTCA 540
 ATCAAGCAGT CCA 553

30

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 869 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

40

GGCAGGAGCT GCCAAGACTG AGGTCTTCGT GGCTTCTCAC ATCTAGATGT ATCCCTCTCA 60
 AATCTATCCT CTATCCAGGC ACCAGATTGA GGTATCTAAA ATGTCAACTT TCCAGTTACT 120
 45 CCTTCTTATA CTAGCCCAAT CAACTTACAA GATAAAGTCC AAGCCCTTC ATATGACAAA 180
 CCACAGCCTG CTAACTCTC CAGGTTTGAA TCCTTCATCT CTTACTTTAA ACTTTAAAAC 240
 CCAGCAGCAC GAAAGTGTCT CCTATGCATG TTGCCATATG CGTTCTCTCC ATCATGCATT 300
 50 TGCCTGAGCA AGATGTCTTG AGTTAACATC TTATTCCTTA AGACTCATTG TGGTGGTAGA 360
 CAGCCTTTAA TAACGGATCC TTGGCCAGGC ACAGTGACTC ACACCTGTAA TCCCAGAACT 420
 55 TTGAAAGGCC AAAGAAGGAA GAAGCTTGA GGCCAGTAGT TTGAGACCAG CCTGGGAAAC 480
 AGAGAGATAT CCCATCTGTA CCAAAAATTT AAAAAATAT TAGCAGGGAG TAGTGGCATG 540
 CACAAGTGST CCCAGCTCCA TGGGAGASTG AGGTAGGAAC ATCACTTGAG CCCAGGAAGT 600

60

276

CAAGGCTGCA GTGAACCATG ATCAGAACAT TGCANTCCAG CTTGGGTAAC AGAGTGAGAC 660
 CTTAGGTCAG AAAAATGAAT AAATAAGCAT AAAATTTTAA AACTTTAGCC AGGCATGGTG 720
 5 GCACACATCT GTGGTCCCTG CTACTTAGGA GGCTGAGGTG AGAGGATCCT TGAGCCCAGG 780
 AGCTCAACAC TACAGTGAGC TATGATTGTG CCACTAAACT CCAACCTGGG TGAAAAAGCA 840
 AAACCCCTGCC AAAAAAAAAA AAAAAAACT 859

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 959 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GCGGAGCCGA GATCGTGCCA TTGCACTCCA GCCTGGGCAA CAAGAGTGAA ACTCTGTCTC 60
 AAAAAAAAAA AATTATAATA CTATATGCCA TAAAATGACA TTTCATATTT AAAGAGTTTT 120
 TTAAAACTCT TGTATTTCACA TGCCATAATT TGAAACCCTA TTTCAGTGAA TGAGAATGGT 180
 30 ATCTGTGTGC CTCATTTTTT CATTTTTATC CTTAACAATT TCCACCACAG CCAGTGCATA 240
 TAATGGCAAT GACACCCAGG GATGGAATGA TAAGTTCCAT CRMGCTCAG TCAAGACGCA 300
 GACTTGATGT GGGCCCAACA ACAGTCAATA ATGGAGTCTC CAAAATAAAG CTCTATAGGA 360
 35 AAGGTAAATA CCGGCTGCAC AAGAAACCAC AGCATCTAGG TTCTAACCCC ATCTCTATGA 420
 AGAGCTTGCT GGGAGAGTTT TGACATTWAA CAATCTGTCT GATKGCCAAT TTTTTCCTTC 480
 40 TATAAAATGA TAATGTTTGA YTCAAAGATC CAAAGTCAAT TCATGGTCTA AAACTTAATG 540
 ATTTTTTTAG GTTTTGKAC ATTTCACTGT ACACTGTAGT AATTATATC TTATTTTCCC 600
 ACTAATTTAG AAAAATATAT AAATGATCCT TAATTGGCAA TGGGTCCCTAA GAATTTTGT 660
 45 TTAATCCCT GTTACCCAAA AGAGCCCTTT TTTGTATCTC GCAGTAGTTA CAAGGATCTT 720
 TCTAAATCTT AAAAAAAAAA AAAAAAGAAA GAAAGAAAAG AAAAGAAAAA AAGTCAGCCC 780
 50 GCGGTGGTGG CTCATGCCCTG TAATCCACAG ACTTTGGGAC CAAGGTGGAC AGATCAGCAG 840
 GTCAGGAGAT GGAGACCATC CCGGCCAACA TGGAGAAACC CTGTCTCTAC TAAAAAAAAA 900
 55 AAAAATCGA GGGGGGCCCC GTACCCAATN CGCCGGCTAG TGGTCGTAAA ACAATCAA 959

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1446 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5	CGGGGCGAGG CTGTGTGGCA CCGCCAGGGA GCGGGCCAC CTGAGTCACT TTATGGGTT	60
10	CAGTCAACAC TTCTTGCTC CCGTTTCTT CTCTGTGGG ATGATCTCAG ATGCAGGGG	120
	TGGTTTGGG GTTTTCTCTC TTGTGCAAG GCGTGGACAC TGCTGGGGG CTGGAAGCC	180
15	CCTCCCTTCC TGTCTTCTG TGGCTCCAT CCGCTCATGG GTGCTGCCAT CCTTCTGGA	240
	GAGAGGAGG TGAAAGCTGG TGTGAGCCCA GTGGGTTCCC GCGCACTCAC CCAGGAGCTG	300
20	GCTGGGGCAG GACCGGGAGA GGGAGCACTG CTGCCCTCCT GCGCTGCTC CTTCGGCAGT	360
	TAGGGGTGGA CCGAGCTTCG CTTCGCCAC TGTCTGGAG GGAAGGGGAA GGAGGGGTC	420
	TTGAGCTCG AGCCAGCTG GCGCTGCTG GTGGAGAGAT GAGATTTAGG GCGTGCCTCA	480
25	TGGGGTGGC AGGCCTGGG TGAAATRAGA AAGGCCGAGA ACGTGCAGGT CTGCGAGGG	540
	GAAGTGTCTT GAGTGAAGGA GGGGACCCCA ATCTTGGGG ATGCTGGGAG TGAGTGAGTG	600
30	AGATGGCTGA GTGAGGTTA TGGGGAGCCT GAGGTTTAT GGGCTGTGT ATCCCTTCT	660
	CCCGGGCCCA GCGTGGCTCC CTCCTGCGG CCGGGCCAC AGGTCTCCT CTGGTCCCTG	720
	TCCCTCTGGT GCTTGGGGAT GGAGCGGCAG CAAGGGGTGT AATGGGGCTG GGTCTGTCT	780
35	TCTACAGGCC ACCCGAGGT CCGAGTGGT TGCTGGGGA GCGGACGGG GCTCCTGAGG	840
	GGTACAGGTT GGGTGGGCC TCCCTGAGG TCTGGGTGA GGCTTTGGT CTGCTGCCTC	900
40	TCAGTCACCA AGTCACCTCC CTCTGAAAAT CCGTCCCTT CTTTGGATGT CTTTGTGAGT	960
	CACTCTGGG CTGGCTGTG TCCCTCCTCA GCTTCTGTT CCGGGACAA GGGTCAAGCC	1020
	AGGATGGGC CAGGCTGGG ATCCCCCACC CAGGACCCC CAGGCCCCCT CCCCTGCTGC	1080
45	TTTGGGGGG GCAGGGCAGA AATGGACTCC TTTGGGTCC CCGAGGTGG GTCCCTCCC	1140
	AGCCCTGCAT CTTCCGTGCC STAGACTGC TCCGAGAGG AGGGGCTTG ACCCACAGGA	1200
50	CGTGTGGTG CCGCTGGAC TCAGGGACCC CCAGCTGCC CAGGCTGGT CTCTGGGCA	1260
	TCTCTTCCCT CTGTCCGA AGATCTGGC CTCTAGTGC TTTGAGGGG TTCCCATAT	1320
	CCCTCCCTGA TATTGTATTG AAAATATTAT GCACACTGTT CATGCTTCTA CTAATCAATA	1380
55	AACGCTTAT TTAAAGCCAA AAAAAAAAAA AAAAAACTG AGGGGGGGC CGTACCCAT	1440
	TGCGCA	1446

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1471 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

5	CAAAAAATAA TAATGATAAT TTRAAATAAA TAAGTAACTA ATRAAAAGAT TTTATATCCC	60
15	AGTCTTATGA TGTTCGTTGG CAAGGCTAGA TAAAAAGATG TTAGAATGAA AGAACATATT	120
	TTTAGTGATA TGTAAATGAA GGATTCTACA ATAGTCATAT ATTTTATAT GAATGAATGT	180
	TGGGTTGGGC TGGAGAGGTA TGTGTGTGTA AATATAAAGG TCTCACATTC AGAGTATAGC	240
20	TCTGAAATAA TGGAACTCAT GTCTACAATT CAACATGCAT CTGTATAGTT ACATCTCATG	300
	TAAATATACA CAGACATATT TTGCAGCCAG TAATTGACAG TTAATGTCCA AAACAGGTGA	360
25	TTGATAGGTA ACAGAAATTA GATAACCACC AATTTTGCCC AAGAGAAAGA CTAGAAGGAC	420
	TAAAAGCAGT TGAATGTATG GTACTGACAT TGTGATAAGC AGTCTGATAA CCAGTTTATT	480
	GAAACGTGTG CATTAAACAGA GAATTTAATT TTAACCCCAT AATTTCTCCT ATCCATTAAA	540
30	ATATTATAAT TGTAGTAGT ATGAAACCAA CAGGAAATGT TTTTAATCA TTTAGTGAGG	600
	TGATTCATTT GTTTCATGGG CAAACACTAT CCAGGAAAAG CCTTGCTTGC CTGTTTCCCA	660
35	AAGAGCTCTA AGAAATAGAA TCAAGTGTAA AATGCTTCAG ACCATTTCAGG ATTTCTTGTC	720
	ACTCTTCTCA ACCCCGATCT TCCTGTTATT ACTGATGTTT GAAACCCTGT CATTAGCCCC	780
	GGCCTGGTTA AAGCCCCCTCA GAGTCACCTC TCATTCATAG CAATAGAATT CAACCCCAAG	840
40	TGGTTGATGG TGTCCCCAGC ACACCCGAGA GACCTGATCT CTGGATTGAG TGCTTTTAGC	900
	TCTTCGAGTT TACCCTAAGA TACCTTCGGG CAATATTTTT AACCAACCCA AAGCTCTTC	960
45	AGGTCAATTC TGAAGAGGAC AAGGTGAATC TTGGCTTGGA ACACCATTTT TGGGCTCTTG	1020
	CTACTGAATG AATCAGAAAG GAATTTTTTC TGAAGAGCAT TAGAAAGTAA AGGAGATGTT	1080
	AAAATAAGTT CTTGAAGTAT GTTTTATATT TATCTAAAAC ACTGATTTTA AAAGTTTACA	1140
50	TTCAAATGEG TATTCAAAG AAGTACTGAT TTGTAATTAT TATAGTTTGT GTGTATCATC	1200
	CCCTTTTAAC CGTGCCTAAC AACTGTACTT AAATTTTGTT TTCTAGTGT AACAAATGTT	1260
55	TCCCTAAGA TTTTCTAGAG CCAAATAATG GGAGTGAAAA ATTCCCTAAG TGTTATATAA	1320
	GAAATATAT TAGAAATCA GCTTTGGATT ATACGATTTT TAAATATAC TAATACAGAA	1380
	TCCTCAGTAA TATGTTTGA ATTGGATTTT TTCTCAGAAC TGTACATAA TAAATAATAC	1440
60	ATCAACCAGA AAAAAAAAAA AAAAAAATTN C	1471

5 (2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1402 base pairs
(B) TYPE: nucleic acid
10 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

15 AGGGACGTCT TGCCCTGAGGA GATGCCCATTT TCTGTCTCTGG RTTACCCCTCA CTGCGTGGTG 60
CATGAGCTGC CAGAGCTGAC GCGCGAGAGT TTGGAAGCAG GTGACAGTAA CCAATTTTGC 120
20 TGGAGGAACC TCTTTTCTTG TATCAATCTG CTTCCGATCT TGAACAAGCT GACAAAGTGG 180
AAGCATTCAA GGACAATGAT GCTGGTGGTG TTCAAGTCAG CCCCCATCTT GAAGCGGGCC 240
CTAAAGGTGA AACAAGCCAT GATGCAGCTC TATGTGCTGA AGCTGCTCAA GGTACAGACC 300
25 AAATACTTGG GCGCGCAGTG GCGAAAGAGC AACATGAAGA CCATGTCTGC CATCTACCAG 360
AAGGTGCGGC ATCGGCTGAA CGACGACTGG GCATACGGCA ATGATCTTGA TGGCCGGCCT 420
TGGGACTTCC AGGCAGAGGA GTGTGCCCTT CGTGCCAACA TTGAACGCTT CAACGCCCCG 480
30 CGCTATGACC GGGCCACAG CAACCCCTGAC TTCCTGCCAG TGGACAAGTG CCTGCAGAGT 540
GTCCTGGGCC AACGGGTGGA CCTCCCTGAG GACTTTCAGA TGAATATGA CCTCTGGTTA 600
35 GAAAGGGAGG TCTTCTCCAA GCGCATTTCC TGGGAAGAGC TGCTGCAGTG AGGCTGTTGG 660
TTAGGGGACT GAAATGGAGA GAAAAGATGA TCTGAAGGTA CCTGTGGGAC TGTCCTAGTT 720
CATTGCTGCA GTGCTCCCAT CCCCCACCAG GTGGCAGCAC AGCCCCACTG TGTCTTCCGC 780
40 AGTCTGTCTT GGGCTTGGGT GAGCCAGCT TGACCTCCCC TTGGTTCCCA GGGTCTCTCT 840
CCGAAGCAGT CATCTCTGCC TGAGATCCAT TCTTCTTTA MTTCGCCCAM CCTCCTCTCT 900
45 TGGATATGGT TGGTTTTGGC TCATTTTACA ATCAGCCCAA GGTGGGAAA GCTGGAATGG 960
GATGGGAACC CCTCCGCCGT GCATCTRAAT TTCAGGGGTC ATGCTGATGC CTCTCGAGAC 1020
ATACAAATCC TTGCCTTTGT CAGCTTGCAA AGGAGGAGAG TTTAGGATTA GGGCCAGGGC 1080
50 CAGAAAGTCC GTATCTTGGT TGTGCTCTGG GGTGGGGGTG GGGTGTCTCT GATGTTATTC 1140
CAGCCTCCTG CTACATTATA TCCAGAAGTA ATTGGGGAGG CTCCTTCAGC TGCCTCAGCA 1200
55 CTTTGATTTT GGACAGGGAC AAGGTAGGAA GAGAAGCTTC CCTTAACCAG AGCGGCCATT 1260
TTTCTTTTGG GCTTTCGAGG GCCTGTAAAT ATCTATATAT AATTCTGTGT GTATTCTGTG 1320
TCATGTTGGG GTTTTAAATG TGATTGTGTA TTCTGTTTAC ATTAAAAAGN AGCAAAAATA 1380
60

ATAAAAAAAA AAAAAAAAAA CT

1402

5

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1047 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGCACAGGGG	ACTACAGCCA	CCCACGACCA	TACCCAGCTA	ATTTTGTAT	TTTTTGTAG	60
AGATGGGGTT	TCACGATGTC	CCCAGGCTG	GTCTTGAAC	CCTGGGCTTG	AGCGATCTTC	120
CCATCTTTCC	ATCTTGGCCT	CCTAAAGTGC	TGGGACTGCA	GGCATGAGCC	ACCATGCCCA	180
GCCAAGATTC	TTATTGATTA	CCATGTGCT	TCAAGAAGCC	AAGCCAGTTT	CCAATATTCC	240
CCATTGCTG	GAGTCTTGGT	ACTTTGGGTA	GAAGCAACTG	GTAAATTGTT	AATTGGAACA	300
NTTGGTGGTG	TAGATAACCA	CGTATGCCCA	AACCTAGAGC	ATCTAGGCTC	ACAATTACTA	360
TCCTGACTTG	ATAACAAGTG	TTCTGATATT	AACCTGAAAA	TGGGAATAAT	GCCAAATCTG	420
TGTAACITAA	CATCTATATA	CACAGTGGGG	AGAACTGAAG	TTATTAAACC	TGGAATCTCT	480
GTGATCAAGG	CTAACAGTAG	TTATCTAAGA	AGCAAAGGAC	CTACAATTCT	TAGACTTGGA	540
GTCATATTCT	TTAAGGACGT	GTTCTGAAAC	TATATCAAGC	ATCTGGTTTC	CACGTATTTT	600
TCCCTCAGAA	ATTATGAAGT	ACAAGTAAAA	ATGAAGGTAC	AGGGTAAGAC	ACATGCTGCT	660
TTCTTGCTCT	TGAGTGGAGA	CAGTTTTC	GCCATCTTAA	CCCCTWACA	CAAAACAATT	720
TGTGTTTTAT	AGCAAATAAG	TGACTCAACA	TAATTTCAAT	ATGATGTTTA	TCCACCAGTA	780
CTTTCCTTTC	AGCTTCTAGT	CCCATAARTG	GTTTGTGAAG	TCATCGGTTA	CATTAGCCAA	840
GATAGGCCTA	GACTTGAAGT	CTAGAATGTT	TTCCCACTA	TATGCCAAAG	TAGAATGTGG	900
GTATCTCAGG	GTCATTTTTG	TTGTTC AATT	TCCCACCTGT	ACAGTTGTTA	TGATTCACCT	960
TCCTTATGTG	TCTAATAAAT	CTTGTTCAT	GAAATGATCA	AAAAAAAAAA	AAAAAAACT	1020
CGAGGGGGGG	CCCGGTACCC	AAATCGC				1047

55

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 990 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

5 TTGGAAGGG TCTAGCTCTT TCTCATTCAC CAACTATATT AGAAGCACTT GAGGGAAATT 60
 TACCACTCCA AATCCAAAGC AATGAACAGT CTTTCTCGGA TGATTTTATT GCCTGTGTCC 120
 CAGGATCAAG TGSTGGAAGG CTTGCAAGGT GGCTTCAGCC AGATTCATAT GCGGATCCTC 180
 10 AGAAAACATC TTTGATCCTG GAATAAGGAT CATATTCGTT GTGGTTGGCC TACCACCATA 240
 ACTGTTCAAA CAAAGACCA GTATGGGGAT GTGGTACATG TTCCCAATAT GAAGGTAATT 300
 15 ATAAGTGGAT TAAATTAGCA GACATCTATA TACTGGCTGC AATGACTGAT AAAATTTTAG 360
 AAATGCCAAG TGCTCAGGT CCATTTGTTT TACCTCTTT ATATAAAGG TGATGCTGAA 420
 AGTTTGTTTA AATGACTTGT TTATATTAAT TAGTCCCAA GTGTCCAAGT TACACCTGTT 480
 20 TTTTGTGTA GTTTGTTCTT TACATTTTGC TACCTGTTAC GGGGACTCAA AGGAGGGATA 540
 AGAAAGTATC CATCTAAAGA GTGCTAGACA CATACAGTGA AGCCCTCPA TATGTATTGA 600
 25 TTGAATRAAT GCATGAAAGA ATACATTTTT AAATTTTGTG TATAGTTTTG AAAGACTCAA 660
 GTACGTTCTG TGTTTGGTAT TACTGAAACC ACATTTTAAA AATAACACTC ATTAAGTTAG 720
 AAATATATGA GTTTAGATTG TAAAAGAATG AGGAATTGAA ATAGTTGTAT ACCATATTGA 780
 30 TGAATATAGA GTTTTAGGA TACCTCTTAC CTGAAATATT AATAATATG TTTNCAGACC 840
 ATATTATACA TAATTATTTG TGATTTAATC TGTTAATATG AATATCTCAT TTAACACTTT 900
 35 TATTTCTGAA AAAATTATAT TGAATAAAT TTTATATAGG CAGTCCCCAG CCCTTTCCTC 960
 CTTCAAAGTT GTCTTATAGA GTGATTGTT 990

40

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

45

- (A) LENGTH: 1208 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

TAATCGCTAC TATAGGGAAA GCTGGTCGCT GCAGGTACCG GTCCGGAATT CCGGGTCGAC 60
 CCACGCTCC GAGCGAAATG GCGCTCCGG CCGCGGCCG GCGCTCCGGC GGCTCCGGG 120
 35 AGGTAGACGA GCTGTTCGAC GTAAAGAAGC CCTTCTACAT CGGCAGCTAC CAGCAGTGCA 180
 TAAACGAGGC GCASGGGTGA ACCTRTCAAG CCCAGAGAGA GACGTGGAGA GGGACGTCTT 240
 60 CCTGTATAGA GCTACCTCG CCGAGAGGAA GTTCGGTGTG GTCTGGATG AGATCAAGCC 300

CTCCTCGGCC CCTGAGCTGC AGGCCCTGCC CATGTTTGCT GACTAGCTGG CCCAGGAGAG 360
 TCGGAGGGAC AGCATCGTGG CCGAGCTGGA CCGAGAGATG AGCAGGACCK TGGACGTGAC 420
 5 CAACACCACC TTCCTGCTCA TGGCCGCTC CATCTATCTC CAGGACTAGA ACCCGGATGC 480
 CCGCCTGGGT GCGCTGCACC AGGGGGACAG CCTGGAGTGC ACAGCCATGA CAGTCCAGAT 540
 10 CCTGCTGAAG CTGGACCGCC TGGACCTGCC CCGGAAGGAG CTGAAGAGAA TGCAGGACCT 600
 GGACGAGGAT GCCACCTCA CCCAGCTGCC CACTGCCTGG GTCAGCCTGG CCACGGGTGG 660
 TGAGAAGCTG CAGGATGCCT ACTACATCTT CCAGGAGATG GCTGACAAGT GCTGCCCCAC 720
 15 CCTGCTGCTG CTCAATGGGC AGGCGGCTG CCACATGGCC CAGGGCCGCT GGGAGGCCGC 780
 TGAGGGCCTG CTGCAGGAGG CGCTAGACAA GGATAGTGGC TACCCGAGA CGCTGGTCAA 840
 20 COTCATGCTC CTGTCCAGC ACCTKGGCAA GCGCCCTGAG GTGACAAACC GATACCTGTC 900
 CCAGCTGAAG GATGCCACA GGTCCCATCC CTTTCATCAAG GAGTACCAGG CCPAGGAGAA 960
 CGACTTTGAC AGGCTGGTGC TACAGTACGC TCCCAGCGCT GAGGCTGGGC CAGAGCTGTC 1020
 25 AGGACCATGA AGCCAGGACA GAGGCCAGGA GCCAGCCCTG CAGCCCTGCC CACCCGGCAT 1080
 CCACCTGCAT CCTCTGCGG CAGGAGCCCA CCCCCAGCAC CCCCATCTGT TAATAATAT 1140
 30 CTCAACTCCA RGGTGTTCCTA CCTGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1200
 AAAAAAAAAA 1208

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1922 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GTGCTGGGCT ACTGAGCAGC GCCATGGAGG ACTCTGAAGC ACTGGGCTTC GAACACATGG 60
 GCCTCGATCC CCGGCTCCTT CAGGCTGTCA CCGATCTGGG CTGGTCGCGA CCTACGCTGA 120
 50 TCCAGGAGAA GGCCATCCCA CTGGCCCTAG AAGGGAAGGA CCTCCTGGCT CGGGCCCCGA 180
 CCGGCTCCGG GAAGACGGCC GCTTATGCTA TTCCGATGCT GCAGCTGTTG CTCCATAGGA 240
 55 AGGCGACAGG TCCGGTGGTA GAACAGGCAG TGAGAGGCGT TGTTCTTGTT CCTACCAAGG 300
 AGCTGGCAGG GCAAGCACAG TCCATGATTC AGCAGCTGGC TACCTACTGT GTCGGGATG 360
 TCCGAGTGGC CAATGCTCA GCTGCTGAAG ACTCAGTCTC TCAGAGAGCT GTGCTGATGG 420
 60

	AGAAGCCAGA TGTGGTAGTA GGGACCCCAT CTGCCATATT AAGCCACTTG CAGCAAGACA	480
	GCCTGAAACT TCGTGACTION CTGGAGCTTT TGGTGGTGGG CGAAGCTGAC CTTCCTTTTTT	540
5	CCTTTGGCTT TGAAGAAGAG CTCAAGAGTC TCCTCTGTCA CTGCCCCGG ATTTACCAGG	600
	CTTTTCTCAT GTCAGCTACT TTAAACGAGG ACGTACAAGC ACTCAAGGAG CTGATATTAC	660
10	ATAACCCGGT TACCCCTAAG TTACAGGAGT CCCAGCTGCC TGGGCCAGAC CAGTTACAGC	720
	AGTTTCAGGT GGTCTGTGAG ACTGAGGAAG ACAAATTCCT CCTGCTGTAT GCCCTGTCTA	780
	AGCTGTCAAT GATTCCGGGG AAGTCTCTGC TCTTTGTCAA CACTCTAGAA CGGAGTTACC	840
15	GGCTACGCT GTTCTTGGAA CAGTTACGCA TCCCCACCTG TGTGCTCAAT GGAGAGCTTC	900
	CACTGGGCTC CAGGTGGCAC ATCATCTCAC AGTTCACCA AGGCTTCTAC GACTGTGTCA	960
20	TAGCAACTGA TGCTGAAGTC CTGGGGGCCC CAGTCAAGGG CAAGCGTCGG GGCCGAGGGC	1020
	CNAAAGGGGA CAAGGCTCTT GATCCGGAAG CAGGTGTGGC CCGGGGCATA GACTTCCACC	1080
	ATGTGTCTGC TGTGCTCAAC TTGATCTTC CCCCACCCC TGAGGCTTAC ATCCATCGAG	1140
25	CTGCCAGGAC AGCAGCGGCT AACCAACCCAG GATAGTCTT AACCTTTGTG CTTCCCACGG	1200
	AGCAGTTCCA CTTAGGCAAG ATTGAGGAGC TTCTCAGTGG AGAGAACAGG GGCCCCATTC	1260
30	TGCTCCCGTA CCAGTTCCGG ATGGAGGAGA TCGAGGCTT CCGCTATCGC TGCAGGGATG	1320
	CCATGCCGTC AGTGAATAAG CAGGCCATTC GGGAGCCAG ATTGAAGGAG ATCAAGGAAG	1380
	AGCTTCTGCA TTCTGAGAAG CTTAAGACAT ACTTTGAAGA CAACCCTAGG GACCTCCAGC	1440
35	TGCTGGGCA TGACCTACCT TTGCACCCCG CAGTGGTGAA GCCCCACCTG GGCCATGTTT	1500
	CTGACTACCT GGTTCCTCCT GCTCTCCGTG GCCTGCTGCG CCTCACAAG AAGCGGAAGA	1560
40	AGCTGTCTTC CTCTTGTAGG AAGGCCAAGA GAGCAAAGTC CCAGAACCCA CTGGCCAGCT	1620
	TCAAGCACAA AGGAAAGAAA TTCAGACCCA CAGCCAAGCC CTCTGAGGT TGTGGGGCT	1680
	CTCTGGAGCT GAGCACATTG TCGAGCACAG GCTTACACCC TTGTTGGACA GGCGAGGCTC	1740
45	TGGTGCTTAC TGCACAGCCT GAACAGACAG TTCTGGGGCC GGCAGTGTG GGCCCTTAG	1800
	CTCTTGGCA CTTCCAAGCT GGCATCTTGC CCTTGACAA CAGAATAAAA ATTTAGCTG	1860
50	CCCCAAAAA AAAAAAAAAA AAAAAAATC GAGGGGGGGC CCGTACCCAA TTGGCCCTAT	1920
	AA	1922

55

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1951 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

5	TCGTCCCCAG AGCGGGCTGA GCGCCAGGCG SAGGGTGGCG GGGGAGCCTG GGGGAGCCCG	60
	CGCCACCTCC ACGGGCCTCT CTGAGCTCGG ACACCAGCGC CCTGTCTTAT GACTCTGTCA	120
10	AGTACACGCT GGTGGTAGAT GAGCATGCAC AGCTGGAGCT GGTGAGCCTG CCGCGTGCCTT	180
	CGGAGACTAC AGTGACGAGA GTGACTCTGC CACCGTCTAT GACAACTGTG CCTCCGTCTC	240
15	CTCGCCCTAT GAGTGGGCGA TCGGAGAGGA ATATGAGGAG GCGCGCGCGC CCCAGCCCCC	300
	TGCCTGCCTC TCGGAGGAAC TCCAGCGCTG ATGAACCCGA CGTCCATTTC TCCAAGAAAT	360
	TCCTGAACGT YTTGATGAGT GCGCGCTCCC GCTCCTCCAG TGCTGAGTCC TTCGGGCTGT	420
20	TCTCCTGCAT CATCAACGGG GAGGAGCAGG AGCAGACCCA CCGGGCCATA TTCAGGTTTG	480
	TGCCTCGACA CGAAGACGAA CTTGAGCTGG AAGTGGATGA CCCTCTGCTA GTGGAGCTCC	540
25	AGGCTGAAGA CTA CTGCTAC GAGGCCTACA ACATGCCGCAC TGGTGGCCCG GGTGTCTTTC	600
	CTGCCTATTA CGCCATCGAG GTCACCAAGG AGCCCCAGCA CATGGCAGCC CTGGCCAAAA	660
	ACAGTGACTG GGTGGACCAG TTCCGGGTGA AGTTCTCTGG CTCAGTCCAG GTTCCCTATC	720
30	ACAAGGGCAA TGACGTCTTC TGTGCTGCTA TGCAAAAGAT TGCCACCACC CGCCGGCTCA	780
	CCGTGCACCT TAACCCGCCC TCCAGCTGTG TCCTGGAGAT CAGCGTGGCG GGTGTGAAGA	840
35	TAGCGGTCAA GCGCGATGAC TCCAGGAGG CCAAGGGGAA TAAATGTAGC CACTTTTTC	900
	AGTTAAAAAA CATCTCTTTC TCGGATATC ATCCAAAGAA CAACAAGTAC TTTGGGTTC	960
	TCACCAAGCA CCGCGCCGAC CACCGGTTTG CCTGCCACGT CTTTGTGTCT GAAGACTCCA	1020
40	CCAAAGCCCT GGCAGAGTCC GTGGGGAGAG CATTCAGCA GTTCTACAG CAGTTTGTGG	1080
	AGTACACCTG CCCCACAGAA GATATCTACC TGGAGTAGCT GTGCAGCCCC GCCCTCTGCG	1140
45	TCCCCCAGCC CTCAGGCCAG TGCCAGGACA GGTGGCTGCT GACAGGATGT GGCACCTGCTT	1200
	GAGGAGGGGC ACCTGCCACC CCCAGAGGAC AAGGAAGTGG GCGGCTGGCC CAGGCTAGGG	1260
	GAGGGTGGGG CAATGGGGAG AGGCAAATGC AGTTTATTGT AATATATGGG ATTAGATTCA	1320
50	TCTATGGAGG GCAGAGTGGG CTGCCCTGGG ATTGGGAGGG ACAGGGCTTG GCGAGCAGGT	1380
	CTCTGGCAGA GAAGGATGTC CGTTCCAGGA GCACACGGCC CTGCCCATC CTGGGCCTTA	1440
55	CCTCCCTGTC CAGGGCTCGG GCGCTGTGGC TCCTGCCTTG ATGAAGCCCG TGTGCTGCCT	1500
	TGATGAAGCC TGTGCCACCT GCAAGTCCCC GCCTGCCCC TGCCCCAACC CCCACCGAAG	1560
	AGCCCTGAGC TCAGGCTGAG CCCAGCCACC TCCCAAGGAC TTTCCAGTGA GGAAATGGCA	1620
60	ACACGTGGAG GTGAAGTCCC TGTCTCTAGC TCCGTCTCTT GCGGGGCTTC TGGGTGGCTC	1680

CTGCCACTGA CCTCACCAGC ATGCTGCCCT GTGGCAGGCC TAGGACCTCA GCGGGGAGG 1740
 AGGAGCTGCC GCAAGGCCCT GTCCAGCAG AAGAGGGAGG CTTCCTGACT GACACAGGCC 1800
 5 AGCCCCATCT TGGTCTGTG ACCCTGCCCC CAACTATTAA AGTGGCATTT CCTGTCAAAA 1860
 AAAAAAAAAA AAAATCGGGG GGGGCCCGGA ANCCATTTC CCCCCAAAAG GGGGTTATA 1920
 10 AAAATTCCTN GCGCTGTTT TAAAAATTC G 1951

15 (2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1989 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

25 GGCACAGGCC GCAGGGNACC TATGGGCGCA TATAGTTGT AATGAACTG TAGTCTCAGT 60
 TGAAGCCTA GACATGAAAT GGTTCAGTGA GCAAGGCTCT ATTCCTAGTC TCCAGCCATG 120
 CCTGTGGAAC CTGACCCRC TCTCAGCACA TTGGACCCAG GCAGATGYAA AAAATTCACA 130
 30 GAACTATGAT TTGGACTCAA GGGTTTGTAG ATTCCTCCT TCATCTAAT TTCAGTGTCT 240
 AAAATTCCTG CATCCRTGAA CGAGCTGGGC ATTGATGAG ACAGGGCTGA ATACTGCAGT 300
 35 TTTCCTCCTA GAAATCATCT GGGGCATTTT CTTCGAACTG ATGGGAACAA TAAGGCATAA 360
 CTGTTTGCAC AAACCTGGGA TAATGATTT TGGGATAACG ATCTACCAGA ATGGGGATAT 420
 TTCAACCTTG GTTCTGAGAT GCAAACCAA GAATATCATG ACCAGCTTTC AGGCCTCCTG 480
 40 AAGTATATCT CTCACATTGT CTTGTTCTCA TGCTGAGGAG CCTGAGATCC CTGTGTGGGG 540
 ATTAGACAGT GCACTGTTAT GGTGTAGGT GAATGGCTT ATTTGTCTG TCCCTGTCTG 600
 45 AATGTATTGC AGGAAYTAAA AAGGACCAAG AAGAGGAAGA AGACCAAGGC CCACCATGCC 660
 CCAGGCTCAG CAGGGAGCTG CTGGAGGTAG TAGAGCCTGA AGTCTTGCAG GACTCACTGG 720
 ATAGATGTTA TTCAACTCCT TCCAGTTGTC TTGAACAGCC TGACTCCTGC CAGCCCTATG 780
 50 GAAGTTCCTT TTATGCATTG GAGGAAAAAC ATGTTGGCTT TTCTCTGAC GTGGGAGAAA 840
 TTGAAAAGAA GGGGAAGGGG AAGAAAAGAA GGGGAAGAAG ATCAAAGAAG GAAAGAAGAA 900
 55 GGGGAAGAAA AGAAGGGGAA GAAGATCAAA ACCCACCCTG CCCCAGGCTC AGCAGGGAGC 960
 TGCTGGATGA GAAAGRGCT GAAGTCTTGC AGGACTCACT GGATAGATGT TATTCAACTC 1020
 60 CTTCACTTGT GTTGAAGTGT GTGACTCATG CCAGCCCTAC AGAAGTGCCT TTTATGTATT 1080

	GGAGCAACAG CATGTTGGCT TGGCTGTTGA CATGGATGAA ATTGAAAAGT ACCAAGAAGT	1140
	GGAAGAAGAC CAAGACCCAT CATGCCCCAG GCTCAGCAGG GAGCTGCTGG ATGAGAAAGA	1200
5	GCCTGAAGTC TTGCAGGACT CACTGGATAG ATGTTATTCG ACTCCTTCAG GTTATCTTGA	1260
	ACTGCCTGAC TTAGGCCACG CCTACAGCAG TGCKGTTTAC TCATTGGAGG AMCAKTACCT	1320
10	TGGCTTKKCT CTTGACCTGG ASAAATTGAA AAGAAGGGGA AGGGCAAPAA AAGAAGGGGA	1380
	AGAAGATCAA AGAAGGAAG AAGAAGGGGA AGAAAAGAAG GGAAGAAGA TCAAAACCCA	1440
	CCATGCCCCA GGCTCAGCAG GGAGCTGCTG GATGAGAAAG GGCTGAAGT CTTGCAGGAC	1500
15	TCACTGGATA GATGTTATTC AACTCCTTCA GGTGTGCTTG AACTGACTGA CTCATGCCAG	1560
	CCCTACAGAA GTGCCCTTTTA YRTATTGGAG CAACAGYGTG TTGGCTTGGC TGTGACATG	1620
20	GATGAAATTG AAAAGTACCA AGAAGTGGAA GAAGACCAAG ACCCATCATG CCCCAGGCTC	1680
	AGCAGGGAGC TGCTGGATGA GAAAGAGCCT GAAGTCTTGC AGGACTCACT CGATAGATGT	1740
	TATTCGACTC CTTGAGGTTA TCTTGAAGTG CCTGACTTAG GCCAGCCCTA CAGCAGTCT	1800
25	GTTTACTCAT TGGAGGAACA GTACCTTGGC TTGGCTCTTG ACGTGGACAG AATTAAAAAG	1860
	GACCAAGAAG AGGAAGAAGA CCAAGGCCCA CCATGCCCA GGCTCAGCAG GGAGCTGCTG	1920
30	GAGGTAGTAG AGCCTGAAGT CTTGCAGGAC TCACTGGATA GATGTTATTC AACTCCTTCC	1980
	AGTTGTCTTG AACAGCCTGA CTCCTGCCAG CCCTATGGAA GTTCCTTTTA TGCATTGGAG	2040
	GAAAAACATG TTGGCTTTTC TCTTGACGTG GGAGAAATTG AAAAGAAGGG GAACGGGAAG	2100
35	AAAAGAAGGG GAAGAAGATC AAMGAAGRAA AGAAGAAGGG GAAGAAAAGA AGGGGAAGAA	2160
	GATCAAAACC CACCATGCCC CAGGCTCAAC GGCCTGCTGA TGGAAAGTGA AGAGCSTGAA	2220
40	GTCTTACAGG ACTCACTGGA TAGATGTTAT TCGACTCCGT CAATGTACTT TGAACCTACCT	2280
	GACTCATTCC AGCACTACAG AAGTGTGTTT TACTCATTTG AGGAACAGCA CATCAGCTTC	2340
	GCCCTTTACG TGGACAATAG GTTTTTTACT TTGACGGTGA CAAGTCTCCA CCTGGTGTTC	2400
45	CAGATGGGAG TCATATTCCC ACAATAAGCA GCCCTTASTA AKCCGAGAGA TGTCACTCCT	2460
	GCAGGCAGGA CCTATAGGCA MGTGAAGATT TGAATGAAAG TACAGTTCCA TTTGGAAGCC	2520
50	CAGACATAGG ATGGGTCACT GGGCATGGCT CTATTCTAT TCTCAAACCA TGCCAGTGGC	2580
	AACCTGTGCT CAGTCTGAAG ACAATGGACC CACGTTAGGT GTGACACGTT CACATAACTG	2640
	TGCAGCACAT GCGGGGACTG ATCAGTCRGA CATTTTAATT TGAACCACGT ATCTCTGGGT	2700
55	AGCTACAAAA TTGCTCAGGG ATTTCATTTT GCAGGCATGT CTCGAGCTT CTATACCTGC	2760
	TCAAGGTCAK TGTCACTTTT GTGTTTAGCT CATCCAAAGG TGTTACCCTG GTTTCAATGA	2820
60	ACCTAACCTC ATTCTTTGTG TCTTCAGTGT TGGCTTGTTC TAGCTGATCC ATCTGTAAAC	2880

CAGGAGGGAT CCTTGGCTGA GGATTGTATT TCAGAACCAC CAACTGCTCT TGACAATTGT 2940
 TAACCCGCTA GRCCTCTTTG GTTAGAGAAG CCAAGTCTCT TCAGCCTCCA ATTGGTGTCA 3000
 5 GTACTTAGGA AGACCACAGC TAGATGGACA AACAGCATTG GGAGGCTTA GCCCTGCTCC 3060
 TCTCRATTCC ATCTGTAGA GAACAGGAGT CAGGAGCCGC TGGCAGGAGA CAGCATGTCA 3120
 CCCAGGACTC TGCCGGTGCA GAATATGAAC AAYGCCATGT TCTTGCAGAA AACGCTTAGC 3180
 10 CTGAGTTTCA TAGGAGGTAA TCACCAGACA ACTGCAGAAT GTRGRCCT GAGCAGGACA 3240
 GCTGACCTGT CTCTTCACA TAGTCCATCT CACCACAAAT CACACAACAA AAAGGAGARG 3300
 15 AGATATTTTG GGTTCAAAAA AAGTAAAAAG ATAATGTAGC TGCATTTCTT TAGTTATTTT 3360
 GARCCCAAAA TATTTCTCA TCTTTTGTG GTTGTCTATG ATGGTGGTCA CATGGACTTG 3420
 TTTATAGAGG ACAGGTCAGC TGTCTGGCTC AGTGATCTAC ATTCTGAAGT TGTCTGAAAA 3480
 20 TGTCTTCATG ATTAATTTCA GCCTAAACGT TTTGCCGGGA ACACTGCAGA GACAATGCTG 3540
 TGAGTTTCCA ACCTYAGCCC ATCTGGGGC AGAGAAGGTC TAGTTTGTCC ATCASCATTA 3600
 25 TCATGATATC AGGACTGGTT ACTTGGTTAA GGAGGGGTCT AGGAGATCTG TCCCTTTTAG 3660
 AGACACCTTA CTTATAATGA AGTATTGGG AGGGTGGTTT TCAAAATTAG AAATGTCCTG 3720
 TATTCRATG ATCATCTCTT AAACATTTTA TCATTTATTA ATCATCCCTG CCTGTGTCTA 3780
 30 TTATTATATT CATATCTCTA CGCTGGAAAC TTTCTGGCTC AATGTTTACT GTGCCTTTGT 3840
 TTTTGCTAGT GTGTGTTGTT GAAAAAATA ACATTCTCTG CCTGAGTTT AATTTTGTG 3900
 35 CAAAGTTATT TTAATCTATA CAATTAAAG CTTTTGCCA TCRAAAAAA AAAAAAATA 3960
 AAAAAAATA AAAAGCGGA CGCGTGGC 3989

40

(2) INFORMATION FOR SEQ ID NO: 29:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3735 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

CTGCTGTTGG CTGGCTGGGC TCCGCAGCAG GCTTGGCCAG CSGCTGACGG GTGGGGGGG 60
 GGGTTTGTGT GAACAGGCAC GCAGCTGCAG ATTTTATCTT GGTAGTGCAN CCTCTCAAA 120
 55 GGTGAAGGA ACTGATGTAA CAGGGATTGA AGAAGTAGTA ATTCCAAAA AGAAAACCTG 180
 GGATAAAGTA GCGGTCTTTC AGGCACTTGC ATCCACAGTA AACAGGGATA CCACAGCTGT 240
 60 GCCTTATGTG TTTCAAGATG ATCTTACCT TATGCCAGCA TCATCTTTGG AATCTGTTG 300

	ATTTTACTG	GCAAAGAAAT	CCGGGGAGAA	TGTGGCCAAG	TTTATTATTA	ATTGATACCC	360
	CAAATATTTT	CAGAAGGACA	TAGCTGAACC	TCATATACCG	TGTTTAATGC	CTGAGTACTT	420
5	TGAACCTCAG	ATCRAAGACA	TAAGTGAAGC	CGCCCTGAAG	GAACGAATTG	AGCTCAGAAA	480
	AGTCARAGCC	TCTGTGGACA	TGTTTGATCA	GCTTTTGCAA	GCAGGAACCA	CTGTGTCTCT	540
10	TGAAACAACA	AATAGTCTCT	TGGATTTWTT	GTGTTACTAT	GGTGACCAGG	AGCCCTCAAC	600
	TGATTACCAT	TTTCAACAAA	CTGGACAGTC	AGAAGCATTG	GAACAGGAAA	ATGATGAGAC	660
	ATCTAGGAGG	AAAGCTGGTC	ATCAGTTTGG	AGTTACATGG	CGAGCAAAAA	ACAACGCTGA	720
15	GAGAATCTTT	TCTCTAATGC	CAGAGAAAAA	TGAACATTCC	TATTGCACAA	TGATCCGAGG	780
	AATGGTGAAG	CACCGAGCTT	ATGAGCAGGC	ATTAACTTGG	TACACTGAGT	TACTAAACAA	840
20	CAGACTCCAT	GCTGATGTAT	ACACATTTAA	TGCATTGATT	GAAGCAACAG	TATGTGCGAT	900
	AAATGAGAAA	TTTGAGGAAA	AATCGAGTAA	AATACTGGAG	CTGCTAAGAC	ACATGGTTGC	960
	ACAGAAGGTG	AAACCAAATC	TTCAGACTTT	TAATACCATT	CTGAAATGTC	TCCGAAGATT	1020
25	TCATGTGTTT	GCAAGATCCG	CAGCCTTACA	GGTTTTACCT	GAAATGAAAG	CCATTGGAAT	1080
	AGAACCTCTG	CTTGCAACAT	ATCACCATAT	TATTCGCCCTG	TTTGATCAAC	CTGGAGACCC	1140
30	TTTAAAGAGA	TCATCCTTCA	TCATTTATGA	TATAATGAAT	GAATTAATGG	GAAAGAGATT	1200
	TTCTCCAAAG	GACCCGGATG	ATGATAAGTT	TTTTCAGTCA	GCCATGAGCA	TATGCTCATC	1260
	TCTCAGAGAT	CTAGAACTTG	CCTACCAAGT	ACATGGCCTT	TTAAAAACCG	GAGACACTG	1320
35	GAAATTCATT	GGACCTGATC	AACATCCTAA	TTTCTATTAT	TCCAAGTTCT	TCGATTTGAT	1380
	TTGTCTAATG	GAACAAATTG	ATGTTACCTT	GAAGTGGTAT	GAGGACCTGA	TACCTTCAGC	1440
40	CTACTTTCCC	CACTCCCAAA	CAATGATACA	TCTTCTCCAA	GCATTGGATG	TGGCCAATCG	1500
	GCTAGAAGTG	ATTCTTAAAA	TTTGGAAGA	TAGTAAAGAA	TATGGTCATA	CTTTCCGCCG	1560
45	TGACCTGACA	GAAGAGATCC	TGATGCTCAT	GGCAAGGGAC	AAGCACCCAC	CAGAGCTTCA	1620
	GGTGGCATT	GCTGACTGTG	CTGCTGATAT	CAATCTGCG	TATGAAAGCC	AACCCATCAG	1680
	ACAGACTGCT	CAGGATTGGC	CAGCCACCTC	TCTCAACTGT	ATAGCTATCC	TCTTTTAAAG	1740
50	GGCTGGGAGA	ACTCAGGAAG	CCTGGAATA	GTTGGGGCTT	TTCAGGAAGC	ATAATAAGAT	1800
	TCCTAGAAGT	GAGTTGCTGA	ATGAGCTTAT	GCACAGTGCA	AAAGTGTCTA	ACAGCCCTTC	1860
	CCAGGCCATT	GAAGTAGTAG	AGCTGGCAAG	TGCTTTCAGC	TTACCTATTT	GTGAGGGCCT	1920
55	CACCCAGAGA	GTAATGAGTG	ATTTTGCAAT	CAACCAGGAA	CAAPAGGAAG	CCCTAAGTAA	1980
	TCTAACTGCA	TTGACCAAGT	ACAGTGATAC	TGACAGCAGC	ACTGACAGCG	ACAGTGACAC	2040
60	CAGTGAAGGC	AAATGAAAGT	CGAGATTCTAG	GAGCAGCAAT	GGTCTCACCA	TAGCTGCTGG	2100

	AATCACACCT GAGAACTGAG ATATACCAAT ATTTAACATT GTTACAAAGA AGAAAAGATA	2160
5	CAGATTTCCT GAATTTGTGA CTGTGAGGTA CAGTCAGTAC ACAGCTGACT TATGTAGATT	2220
	TAAGCTGCTA ATATGCTACT TAACCATCTA TTAATGCACC ATTAAAGGCT TAGCATTTAA	2280
	GTAGCAACAT TCGGGTTTTC AGACACATCG TGAGGTCCAT GGCTCTTGTC ATCAGGATAA	2340
10	GCCTGCACAC CTAGAGTCTC GGTGAGCTGA CCTCAGCATG CTGTCTCTGT GCGATTGCCC	2400
	TCTCTGTGTG CTGGACTTCT GCGTTTGTG GCGTGATGTG CTGCTGTGAT GCTGGTCTTT	2460
15	CATCTTAGGT GTTCATGCAG TTCTAACACA GTTGGGGTTG GGTCAATAGT TTCCCAATTT	2520
	CAGGATATTT CGATGTCAGA AATAACGCAT CTTAGGAATG ACTAAACAAG ATAATGGCAG	2580
	TTTAGGCTGC ACAACTGGTA AAATGACTGT AGATAAATGT TGTAAATAGT GTACACGTTT	2640
20	GTATTTTGTG TAATATAGCC GCTGCCATAG TTTTCTAACT TGAACAGCCA TGAATGTTTC	2700
	ATGTCTCCCT TTTTTTTTTC TCTATAGCTG TTACCTATTT TAGTGGTTGA AATGAGAGCT	2760
25	AGTGATGACA GAAGGATGTG GAATGCTCTC TTGACATCAT TGTGTATTGC TGTAATCAA	2820
	GTGGTAACG ACTACTTCTA GCAGCTCTTA CCACTATGAC TTAAGTGGTC CTGGAAGGCA	2880
	GTAAGTGGAG GTTTCAGCA TTCTGCCTTT CATGAGGGCT TCTACCACTG ACCACTTTGC	2940
30	ACGTACCTGG CTCCCAGATT TACTTAGGTA CCCCACGAGT CGTCCACATA ACCAGCTTCA	3000
	TCTTTACCTT GCCAGAGTTG ACAATTATGG GATACTCTAG TCTACTTATA CTTGTGTTCC	3060
35	CATCTGTCTG CCATCCTCTG AAGGCCAGGA CCCAGTCATA CATCCTTAGA AACCAAGTA	3120
	TGGTTTTTGT TTTCTCTTGG AATGTCAGGT CTTAAGGCAT TTAATTGAGG GACAAAAAAA	3180
	AAAAAAGCC GATATAGTAG CTAGCTACTT AAGCATCCAT GGGTATTGCT CCATATCAAA	3240
40	GCAGATTTGC AGGACAGAAA GAGTAAATTA GCCTTCAGTC TTGGTTTACA GCTTCCAAGG	3300
	AGAGCCTTGG CCACCTGAAA TGTAACTCG GTCCCTTCCT GTCTCTAGTT CATCAGCACC	3360
45	TGCAGATGCC TGACTCTTGT TAGCCTTACT ATTCAATACA GTCCTTAGAT TCACGGTATG	3420
	CCTCTTCCTA TCCAGGCACC TATTCTGAAT CACCATGTTG CTCTGCAGCT AGAGTTGATA	3480
	GGAGAAAATC CATTTGGGTA GATGGCCTAT GAATTTGTAG TAGACTTTCA AAATGAGTGA	3540
50	TTTGTTAGCT TGGTACTTTT AAGTTTGTGG TACAGATCCT CCAAACCCAT ACTCTGAGCA	3600
	ATTAAGTCCC TTGAACATAG AGAAAATTAA GGCCTCACAG GATGAGTCTC CATTCTCTGT	3660
55	AAATGCTTAT TTTATCATAG TCTTTAGCCN CTAATATGAG TAAAATGTTT TCTTNGCCG	3720
	GGTGTGGTGA CTCAC	3735

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1667 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

5	TAGTAATTCA TTAACTCCT CTTACATGAG TAGCGACAAT GACTCAGATA TCGAAGATGA	60
	AGACTTAAAG TTAGAGCTGC GAGCACTACG AGATAAACAT CTCAAAGAGA TTCAGGACCT	120
15	GCAGAGTCCG CAGAAGCATG AAATTGAATC TTTGTATACC AAATGGGCA AGGTGCCCCC	180
	TGCTGTTATT ATTCCCCCAG CTGCTCCCTT TTCAGGGAGA AGACGACGAC CCACTAAAAG	240
20	CAAAGGCAGC AAATCTAGTC GAAGCAGTTC CTTGGGGAAT AAAAGCCCCC AGCTTTCAGG	300
	TAACCTGTCT GGTGAGAGTG CAGCTTCAGT CTTGCACCCC CAGCAGACCC TCCACCCCTCC	360
	TGGCAACATC CCAGAGTCCG GGCAGAATCA GCTGTTACAG CCCCTTAAGC CATCTCCCTC	420
25	CAGTGACAAC CTCTATTGAG CTTTCACCAG TGATGGTGCC ATTTCACTAC CAAGCCTTTC	480
	TGCTCCAGGT CAAGGAACCA GCAGCACAAA CACTGTTGGG GCAACAGTGA ACAGCCAAGC	540
30	CGCCCAAGCT CAGCCTCCTG CCATGACGTC CAGCAGGAAG GGCACATTCA CAGATGACTT	600
	GCACAAGTTG GTAGACAATT GGGCCCGAGA TGCCATGAAT CTCTCAGGCA GGAGAGGAAG	660
	CAAAGGGCAC ATGAATTATG AGGGCCCTGG AATGGCAAGG AAGTTCTCTG CACCTGGGCA	720
35	ACTGTGCATC TCCATGACCT CGAACCTGGG TGGCTCTGCC CCCATCTCTG CAGCATCAGC	780
	TACCTCTCTA GGTCACTTCA CCAAGTCTAT GTGCCCCCCA CAGCAGTATG GCTTTCCAGC	840
40	TACCCCATTT GGGCTCAAT GGAGTGGGAC GGGTGGCCCA GCACCACAGC CACTTGGCCA	900
	GTTCCAACCT GTGGGAAGTG CTTCTTTGCA GAATTTCAAC ATCAGCAATT TGCAGAAATC	960
	CATCAGCAAC CCCCCAGGCT CCAACCTGCG GACCACTTAG ACCTAGAGAC ATTAAGTAA	1020
45	TAGATCTGGG GGCAGGAGAT GGAATGCTGA GGGGTGGGT GGGGTGGGA AGTAGCCTAT	1080
	ATACTAACTA CTAGTGCTGC ATTTAACTGG TTATTTCTTG CCAGAGGGGA ATGTTTTTAA	1140
50	TACTGCATTG AGCCCTCAGA ATGGAGAGTC TCCCCGCTC CAGTTATTGG AATGGGAGAG	1200
	GAAGGAAAGA ACAGCTTTTT TGTCAAGGGG CAGCTTCAGA CCATGCTTTC CTGTTTATCT	1260
	ATACTCAGTA ATGAGGATGA GGGCTAGGAA AGTCTTGTTC ATAAGGAAGC TGGAGAACTC	1320
55	AATGTAAAAT CAAACCCATC TGTAATTTGG AGTGGGTGGA GCTCTTGCTT TTGGTACATG	1380
	CCCTGAATCC CTCCTCCCT CAAGAATCCG AACCACAGGA CAAAACCAC CTACTGGGCT	1440
60	CTCTCTACC CTGCCCTCCT CCGTTTTTTC TACCCCTCTC TTTTTATT TTTCTTTGCT	1500

CTTTAGAACT CAGTGA AAAA TACCAGGTA CTGGGTGCA ACTCTTTCTT ATGATAGGTC 1560
 ATTAGTCTT TAAGCAAAAAG ATATTAGCAG CTTTGA CTGC AGCATTAGCA ATTAGGAAAA 1620
 5 AAAAAANWA AAAACTCGAG GGGGGGGGGG GTTACCCAAT TGGCCCT 1667

10 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1408 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

20 ATTACACACC TGAGCACTGT GCCTGGCAAG ACCTGTCTTA ATAGATTAGA GAACCACTGA 60
 TAGATGGTCA GCTTTCTGTA GCAGTGAGAA CCTACATTT CAAATGTGGA TAGCACCTTT 120
 GCGGGGAAAC ATCACTTGGC ACATCTGCAT TCTTTTGTGA CACAGGGTCT CACTCTGTTG 180
 25 CCCAGGCTAG AGTGCAATGGC AGGATCTTAG CTCACTGCCA CCTCCACCTC CCAAGTTCAA 240
 GCGATTCTTC TGCTCAGCC TCGTGAGCAG CTGGGATCAC AGACATGGCC TACCATGCC 300
 30 AGCTAATTTT TTGTATTTT TGTGTGTTG TTTTGTPTK TAAGTAGAGA CCGGCTTTCA 360
 CCACGTTGGS CAGGCAGGTC TCGAACTCCT GAMCTCAGGT GATCCACCCA CATCTGCGTT 420
 CCAATATCTT TCTCAACATA ATGATAGCCG TAATTAATAT TTTCCAGTAC ATTTTATG 480
 35 CTTTACACAC GAGAGTGGA GACAGACACA AACCCAGATC TGTCTGACTC CAAAGCCCGT 540
 TTGTATCAT TCTTTTACG GTATCCTATA GTGGTATCCT TTACAGAAAG ACAGCTTTTA 600
 40 CCCAACAAAG ACTTAACTTC CCAGGATGCC AGAAGGACAA AGCGGGATTG CTTTAAAGRA 660
 GRAAGTTATC AAGAMCTTAT TTTATAAATG AGATTAGATA GGGAAAGGCA ATTTATCTTT 720
 45 ATTA AAAACT GAAAAGGCCA GCATAGGGAA GGAGGTCCTT CGGTGGTCTT TTTCAGGGAA 780
 ATACTTCAGT TGCTTTTAT AGAAACAGAT AGTACCTAAG GTTTTGAGGT AGGWACAGCT 840
 TAAGGCATGC TAATGKTCAT GGGTCCTTCC ATAGTCATTT TGTATTTTG GTTWACATTT 900
 50 GAGCAATAGC CAGCCCTTCA CTGCTGCTGG AYTCACTCCT GCCATTTA CAGGTGACAG 960
 AGGAGACAGG AGGTATGCT TTTCTATTTT TAWACATGCT TTATATTTAA CACAAGCTCT 1020
 TGGTATCTT AGATAAACAG AAGTTGCCTA GCACTCCTTT TAGTGCAATT AACCTTTAA 1080
 55 CATTTAAGCA AAATAATAAA CAGTCTTTTG AGGTTCTTAA ACAATGAAAC GTGTTGAGT 1140
 GGCAGCAGCG GAATCCATGC TCTTCTCCT GGAGTGTGCA AKAGTCCGTG GTCCTGAGTA 1200
 60 TCTCACACAG ATGTGGCAAT TTATGTGTGA TGCTCTAATT AAGGCCATTG GTACAGAACC 1260

AGATTTCAGAC GTCTCTCTGAG AATAATGCA TTCTTTTGCA AAGGTGAATA TTTTCTCTTT 1320
 AAAAAATGTG TATAGGTGG TATGTGATT TATTAGTCTT GCTAAAAAA AAAAAAAA 1380
 ACTTACAGGG GGGGCTGGT ACCCAATT 1408

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2031 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

AGGATATGCA TGAATCTTAA CCAGGCTATA TGTTAAAAA AAATTGGAAA ATGCAATACA 60
 TTTTTRACTA TCAAACTAC AGATGAGTA TGCAAGTTTT ATTTATCAA ATGTAATGGA 120
 TTTTAAAGG CTGAGAAATT TTCTTATAC CTACCTTTTC AGTTATTTTA ATTATACCAA 180
 ATTATCAACT AGAATAGCTT CTTCCATATG AATATAAAA TGAAGAGACA CCTACGCTCT 240
 ATCAGGCTTA GGAATCTTTC AACTTATTTT CACTTTAATT TCTCAGTGGG AGTTAAGAGG 300
 GGTGAGAAA CAAAGAGGG GAAAACTGA CAACTAACAA AACCAGCACC ACATCGCTAG 360
 GTGGTGCTA CTAATTACCT TCTCAGGATT TTCTCAGAT TGAAAAGCTT ATGAGGATTT 420
 CTGCGGATC TTAATAAAGT GCTGTAGT ACAGAGCTTT CCTGATGATA TTTACTCTTG 480
 AGCATATGTG GTTGTAAAC CTTAACTTTC TTTCTCCAGG AGGGTGGTGA TAGAAACAGA 540
 TGGTAGTATT TACGAAGTGA TGTCTCTGTG AATGTTGAG GGTGGGGAGA AAGACTTTA 600
 AGGAGGAGA GCCATCTATT TTGTTCTTAA AGCCACCTCT CAGCAGAATC GTCATGTTTT 660
 TCTGATGCAC CGCTCTGCTT CATGCCAAG ATGACTTGGG AGGCAATCTC AGGAGCTGTG 720
 GACTTAACCT TTGCAAGCA CACTGTCTTT CTCAGCGTTC TCTGCAAGTC AGTAGGTGTT 780
 AGTATGGTTG CAAAGTTCAC TGTCTCAGCA AAGTTGAAGT GGGCTACCTC TCTACAGCTG 840
 TTTCTCTGA GGGAAAAATC TTGAGACCAG ATGGTGGAGC TCTGGAGTCA GAGGAAATGG 900
 GTGTCTTCAG CACAAAGCTG CTGCTTTTAC TTCAGCCACT TCTGACATTT TTACATACCG 960
 AGCCTGAGAT TTGTGATTA TCTCAATCA AATCACTTTG ATGGAGATAA ATATCAAAA 1020
 CTGTTTTATA GTCATTGATT TGGTGAAC AGTAATGGA AATGGTGTG AAGGACTTCT 1080
 CATTTTGGG GCTTTCTTTC CAGAGTCTGT GCTGATGGT GTTCCCTGTT CATCTGAGCC 1140
 CCGAAAGCA TTATTACTGA TACTTCACA CAGTCAAAAG GGCAGACTCG ATGGATGGTC 1200

TTTTATAAGG CATTTAAGGG TACACTACTG TGTTCCTCTG ACCATACATT TTTCTTAGCC 1260
 CCTCAAGTAA TATAGCAGAG AGTTATGAAT GACAATTCCC CTAACCATTC CTCTTCATAT 1320
 5 CTGCTCTCTC CCCTTACCAT CGTAATTCTC CAACTGGTC ATAAAGGCAC TCTGTGAAGA 1380
 TATGCGGGAC TGACATCTTA AGCTCTCACC TGGCTGCGGT AGGAAAGGCC AACTGACGA 1440
 CAAAAAAAAA ATTCTTTTATA AAGATGATAT GGTAAACATGT ATCTTTGCCC TGGGTCTGGG 1500
 10 TGGGTCCAGT CAGTCTCAGA TTTACAAGCA TTTAGGAGCC TAGGTAAAAG CTGCTAGTAT 1560
 TCTTTTAAA GTTACATTTA TGACTTGCAA TGATAGAAA CTCCTTCGAA TTAAATGGCA 1620
 15 TTTTATAATA TTATCTGTGT ACTTCACAGT GTTAAAAATA CCCTCATACG TTATTGCATT 1680
 TGATCTTCAC AGAAAGTGCA TTTTAACCG TACTCTGGGT GCAATAAATA ATATGTAGAA 1740
 ATTTAAGTCC TCCAATTCCA GCATATCCAG TGAGTTTGA CAGTGTGTTT ATGTGGAATG 1800
 20 TTTAAGGATA TACAATTGTA CTTTATATAA ATTGGTTCTT GTTCTTCTTA AATGTGACAT 1860
 GAATAAATTG TGCTGCTACA TTATACTGGA AATTAAACAG GGAAGGGGA AGAGCTCTTG 1920
 25 GCTCCCTTGA GGTCTGCTA GTGGTGTAG GAGTGGTTAC AACTGAGCTT TTAGTAACCA 1980
 TTTAACCCTA TGTAACCTG GTTCTTAATT AAAAAAAT TTCTTTTCC A 2031

30

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
 35 (A) LENGTH: 971 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCGTCGGAA CTGGGCGCGG GCACATCCAC GGGGCGCGAG TGACACGCGG GAGGGAGAGC 60
 AGTGTCTGTC TGGAGCCGAT GCCAAAAACC ATGCATTTCT TATTGAGATT CATTGTTTTT 120
 45 TTTATCTGT GGGGCTTTT TACTGCTCAG AGACAAAAGA AAGAGGAGAG CACCGAAGAA 180
 GTGAAAATAG AAGTTTTCGA TCCTCCAGAA AACTGCTCTA AGACAAGCAA GAAGGGAGAC 240
 50 CTAATAAATG CCCATATGA CGGTACCTG GCTAAAGACG GCTCGAATT CTAATGCGAC 300
 CGGACACAAA ATGAAGGCCA CCCCAATGG TTTGTTCTTG GTGTTGGCA AGTCATAAAA 360
 GGCCTAGACA TTGCTATGAC AGATATGTGC CCTGGAGAAA AGCGAAAAGT AGTTATACCC 420
 55 CCTTCATTTG CATACGGAA GGAAGGCTAT GCACAAGGCA AGATTCCACC GGATGCTACA 480
 TTGATTTTTG AGATTGAACT TTATGCTGTG ACCAAAGGAC CACGGAGCAT TGAGACATTT 540
 60 AAACAAATAG ACATGGACAA TGACAGGCAG CTCTTAAAG CCGAGATAAA CCTCTACTTG 600

CAAAGGGAAT TTGAAAAACA TGAGAAGCCA CGTGACAAGT CATATCAGGA TGCAGTTTTA 660
 GAAGATATTT TTAAGAAGAA TGACCATGAT GGTGATGGCT TCATTCTCC CAAGGAATAC 720
 AATGTATACC AACACGATGA ACTATAGCAT ATTTGTATTT CTACTTTTTT TTTTGTAGTA 780
 TTTACTGTAC TTTATGTATA AAACAAAGTC ACTTTTCTCC AAGTTGTATT TGCTATTTTT 840
 CCCCTATGAG AAGATATTTT GATCTCCCA ATACATTGAT TTTGGTATAA TAAATGTGAG 900
 GCTGTTTTGC AAACCTAAAA AAAAAWAAA AAAACTSGAG GGGGGCCCGT ACCCAANTCG 960
 CCGNATATGA T 971

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1792 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GAACCCCTTT TCTCTGGTA AAGGGTAAGG GGGGGATAA TGTTTACCAC AGGTACGAAA 60
 TAGTCACTTT AACATTGAGA CCTCTGCCCTC ATTGAATTCA GGTTTTTTAA GTACTTGAAA 120
 CTCTTCAGAT TCTCCTTATT TTAGTTTCTT TTTACATTTA TGAAGTAGAA AGCATTGTTT 180
 TGTAAGTGT TTTGAAAATA AATAGCCTAG TCTCTTATCC TCTTTAGCGT GGATTAAAGG 240
 TGAAGTTCTG CAAATGGGAG AGTGTTTACA GTAGATAGCT CAGATTGATT GAACACATTT 300
 GAGGAAGAGA CTCCTGCATG AGATACCAGC ATTTTACAA ATACTTTTTA TGTACATTCT 360
 TTATTTTGTG ATTTTGTCAA CCGTCTCCCC AAGCACATCT TCTTCTCTT TACTATGTCT 420
 ATGTAGGGAA AAACAAAACA AAAAATTGCA CTTACGTTAC ACTCCCAAAA TGTGGGTAAT 480
 CCGTGTCTTT CAAAAACAT TTCTGTTTTT TGTMTTGT TTGGTCAGTCC ATTGCATAAG 540
 TGACAAGTTT GGGTCTTGT GGCACGTATG TATGAAGCGG GAGGGGGATG ASAATTGCCT 600
 GTCCTTCAGT ARGCTGTAAA AGTAATTTAC ATGTAAGTAA AAAGGGAAAA TAGAATAGAT 660
 GCCAAAGTCA TTTATTCAGT CTTAGTTTTT CTTATGTGGC ATTACTGCAT CTGCTAGTTA 720
 GTGAGAAAGC ACCCTCAGCT TTTACTGCTC CCGTCCCTGC CTGCCAACAC ACTTGATGTG 780
 TGCAACAGC CTTCAAGTAT CTGTCAATG ACCTATATAA GGTATTGAAT AAGGTATTCT 840
 GTGCAATTTA GAAATGGACT GGATAAACT TACTTGGTTG TCATTATTTT ATCTCATTTG 900
 TCCTGTTACA TGGCCTATGT TAAGATAATT ATATTGCCAC TAATAATCAA GATGCTAAT 960

	GAGTATTACA ACTGGCTAAT ATCATTTTTT ATATACAAGG GTATGTGTAT ATTTGGAATT	1020
	GRTATGAGAA ACTCATTTGT ACCCATTTGA GTGATATTGC ACAACAAACA CAGATAYCTA	1080
5	CAGACTCCGT TTTCATTTTC TCGTGTCTT TATGATAATG ATCTTTGTAG ATTGCTTATT	1140
	TCTGTACTTT ATCTGTAATA AACTTTGTAG ATCCTGTGAA CCATTACTTT GCCTAAATCA	1200
10	CTTGAGACTT GAGTCTTTAA TAACAAAGCA TCAATATTCA CTAAAGTCAA TCTCTTTTGA	1260
	GTTTCTGTGA CTGGCTAGA AGCTCTTGAC ACTAAGGGAT TAGTGTTAAT TTTCCCTGGG	1320
	GGTGTCCAC TAGGGCATT CTGTATAATG ACTTGATGTT GCCACATAGA CTTCAAGATA	1380
15	TATAATATTT TGAGGATTTT GTTGATTGGC CTATGTTTTA TTGCATAGTG TGAAACGTGT	1440
	AAAGCTTGGT TAACCTGTAT ATAGATAGCT TATTGTTGAC TAGTTATAGT GTATTTAGGG	1500
20	TTGCCTGTAA TATTTAAGCT TCTTACTGA TGTGTGTGCT GGTAGGAACA TATAATTTTT	1560
	GTACATTATA TTTACTGAGA TGTGCCTTT TTTATTTTAC AAATACTTTG GAATTCCAAT	1620
	GTGTTTTTTG CTCCGTGAG GATTAATTTG GAAAGGTTTT TAATGACATT CCACTGATTT	1680
25	CAGATTTTGC TTGAGATTGA CTTCAATAAA TTGTCTGTGA TGTTCACAAA AAAAATTAAA	1740
	AAACTCGAGG GGGCCCCGGT ACCCAANNCG CCGGATATGA TCGTAAACRA TC	1792

30

(2) INFORMATION FOR SEQ ID NO: 35:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 896 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

	AGTTGNANAC AACAGGACCT GAGTCCCTTG GCAGCACCAG TAGGTTGCCC C/TGCYTC/T	60
	GCCAGCTCA C/TGCCACYT T/TGCCCTY TCGGGATGCC TTCCGAGACA GAGYTYTTGC	120
45	CTGCCTGTGG TGGCCAYTCT TTGCTTTTGG TTYTCTTGCC CCTTGGCCTC CCTTTTTGTC	180
	CCCGGGCAGC CTGTGTGTAC CTGCCCTTT CCCTCCCTTC CTTTCCAGGA CAAGCAGGCC	240
50	GAGGAGGTGC GGAAAAACAA GGACCTGAAG GAAGAGGCCT CCAGGTAAAG CCTAGAGGCC	300
	AAAGAACTTT CCAGGTCAGC CGGACAGCTC CAGCAGCTCC ACGTTCCAGG CAGCCTCGMC	360
55	CGCCGGCTGC GCTCCAGCA CTGGGTTTG GGGGAGGGG GGTGGCCAG GGGCGTTTCC	420
	TCTGCTTTTG GTGTTTGTAC ATGTTAAGAA TTGACCAGTG AAGCCATCCT ATTTGTTTCC	480
	GGGAACAAT GACGGGGTGG GARAGGGGAG AGGAGAGAGT TTGGGAAGG GAGATGGAGA	540
60	AGAACTCAG GACATTGCA CCCTGCCCGG CGCAGATCTG ATTTTCACAT CTCTACCTGG	600

ACATTGAGCC TTCCAGGCAC CMTGTTGAGG AGAGATGAAA ACCAGGGCGG TAGPACTTCA 660
 GGGTGAACGA CAGAGTCTTG GGTGGGGCAG CGGCTGCAGG GGGCACCAGA GAACCCAGCC 720
 5 AGAGGGGGTG TGAGTACCAG TGGTGTGCT TCCACCTGC AGCAGGTGGG ATGAGGTCTG 780
 TGTGTGTGTG TGAACCATCA TTTTGTGATC ATCATGACCA ATGAAACATT GAAAAAAXAAA 840
 10 AAAAAAAGTG GAGGGGGGCC CGTACCCAAN TCGCGGNATA GTGATCGTAA ACAATC 896

15 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 912 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

25 TCGACCCACG CGTCCGGTCA GCGAGTCCGA TCCAGCCATG ACAGCCTTCT GCTCCCTGCT 60
 CCTGCAAGCG CAGAGCCTCC TACCCAGGAC CATGGCAGCC CCCCAGGACA GCCTCAGACC 120
 AGGGGAGGAA GACGAAGGGA TGCAGCTGCT ACAGACAAAG GACTCCATGG CCAAGGGAGC 180
 30 TAGGCCCCGG GCCACCCGGG GCAGGGCTCG CTGGGGTCTG GCCTACAGGC TGCTGCACAA 240
 CCCAACCTTG CAGGTCTTCC GCAAGACGGC CCTGTGGGT GCCAATGGTG CCCAGCCCTG 300
 35 ARGGCAGGGA AKGTCAACCC ACCTGCCCAT CTGTGCTGAG GCATGTTCTT GCCTACCATC 360
 CTCTCCCTTC CCGGGCTCTC CTCCAGCAT CACACCAGCC ATGCAGCCAG CAGGTCTCTC 420
 GGATCACYGT GGTTKGGTGG AGGTCTCTCT GCACTGGGAG CCTCARGAG GCTCTGCTCC 480
 40 ACCCACTTGG CTATGGGAGA GCCAGCAGGG GTTCTGGAGA AAAAAAGTGG TGGTTAGGG 540
 CCTTGGTCCA GGAGCCAGTT GAGCCAGGGC AGCCACATCC AGGCGTCTCC CTACCCTGGC 600
 45 TCTGCCATCA GCCTTGAAGG GCCTCGATGA AGCCTTCTCT GGAACCACTC CAGCCCAGCT 660
 CCACCTCAGC CTTGGCCCTC ACGCTGTGGA AGCAGCCAAG GCACTTCTTC ACCCTTTCAG 720
 CGGCACGGAC CT/TTTGGGG AGTGGCCGGA AAGCTCCCSG GCCT/TTGGCC TGCAGGGCAG 780
 50 CCCAAGTCAT GACTCAGACC AGTCCCACA CTGAGCTGCC CACACTCGAG AGCCAGATAT 840
 TTTTGTAGTT TTTATKCTT TGGCTATTAT GAAAGAGGTT AGTGTGTTCC CTGCAATAAA 900
 55 CTTGTTCTG AG 912

60 (2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1382 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

10	AATTCCGCAC GAGCGGAGGC GAGGCAAACT RAGCGCGAAA GTTGTGTGTC GTGTTGGCAG	50
	GAGGCCCTAG AAGCGAAGA CTGTCTAGTC GGACAATGTC ATATTATAAA TTTGGAATGC	120
15	TGAATAGAAA ATTATAGATT TTGATATTGA AGGAAATGAA GCGAAGCTTA AATGAAAATT	180
	CAGCTCGAAG TACAGCAGGC TGTTCGCTG TTCCCTTGTT CAATCAGAAA AACAGGAACA	240
	GACAGCCATT AACTTCTAAT CCACTTAAAG ATGATTCCAG TATCAGTACC CCTTCTGACA	300
20	ATTATGATTT TCCTCCTCTA CCTACAGATT GGGCCTGGGA AGCTGTGAAT CCAGAGTTKG	360
	CTCCTGTAAT GAAAACAGTG GACACCGGGC AAATACCACA TTCAGTTTCT CGTCCTCTGA	420
	GAAGTCAAGA TTCTGTCTTT AACTCTATTC AATCAAATAC TGGAAAGAAGC CAGGGTGGTT	480
25	GGAGCTACAG AGATGGTAAC AAAAATACCA GCTTGAAAAC TTGGRATAAA AATGATTTTA	540
	AGCCTCAATG TAAACGAACA AACTTAGTGG CAAATGATGG AAAAAATTCT TGTCCAATGA	600
30	GTTCCGGGAGC TCAACAACAA AAACAATTAA GAACACCTGA ACCTCCTAAC TTATCTCGCA	660
	ACAAAGAAAC CGAGCTACTC AGACAAACAC ATTCATCAAA AATATCTGGC TGCACAATGA	720
	GAGGGCTAGA CAAAACAGT GCACTACAGA CACTTAAGCC CAATTTTCAA CAAAATCAAT	780
35	ATAAGANACA AATGTTGGAT GATATTCCAG AAGACAACAC CCTGAAGGAA ACCTCATTGT	840
	ATCAGTTACA GTTTAAGGAA AAAGCTAGTT CTTAAGAAT TATTTCTGCA GTTATTGAAA	900
40	GCATGAAGTA TTGGCGTGAA CATGCACAGA AAAGTGTACT TCTTTTGA GATTAGCTG	960
	TTCTTGATTC AGCTGTTACA CTTGGCCCAT ATTATTCGAA GACTTTTCTT ATGAGGGATG	1020
	GGAAAAATAC TCTGCCTTGT GTCTTTTATG AAATCGATCG TGAACCTCCG AGACTGATTA	1080
45	GAGGCGGAGT TCATAGATGT GTTGGCAACT ATGACCAGAA AAAGAACATT TTCCAATGTG	1140
	TTTCTGTGAG ACCGGCGTCT GTTTCTGAGC AAAAACTTT CCAGGCATTT GTCAAAATTG	1200
50	CAGATGTTGA GATGCAGTAT TATATTAATG TGATGAATGA AACTTAAGTA GTGATAAAAG	1260
	GAAGTTTAGC ATAAATTATA GCAGTTTTCT GTTATTGCTT AATTTACCAT CTCCATAGTT	1320
55	TTATAGCTAC TATTGTATTT CACTGTGTGA APTAAAGTAT TTGAATTCTT TAAAAAATA	1380
	AA	1382

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GGGCTACTTC AAAGCCCTGG GCCTTATTTTC TTCAGGTAAA AAAATATAAA GTCAGATCTC 60
ATCCCGGGTG GCCATGCTGT TAGACCCTTT CATCCTTCTC TTCTGCCCTCT TCTCAACAGC 120
TGCCCCAGTCC TGTTTGGAAT TCATATACAT ACAGTTCTAA TACTGATGTA TTTACCCCTCA 180
TAAGCCACTC AACCCAGAAT CTTATTTGAA TTATAATCCA GAAACATCAG GTGACGTGTG 240
AGACTACTGT ATGAGAAAGA GACAGTTTAA GGGTCAGTCC AATGGAAAAA AGAGTTTCTCA 300
GAGCTTTCTT TAGCTTATTC TCATCAAAGA GCTTCTCTCG CAGAAGGAAC CTACTGGTTC 360
CTCCTTTTCCA GTCCTAGAAA TCCTGACCTA GAGTGGCTTA ATCCTGCTAG CACCTCTCTC 420
TCGCACTCTG GTGCCAAATG ACTCCAGGAA CTGGGCCATG ATGTGGTGGG AATGACCTTA 480
CCCTGAGCAT GTCACTCATG CATTGAACAA CAGCTAAGAG CAGAGCTTAG AGCTTAGAGC 540
TGGGCCCTGT AAGGTGAGAG GAATCAGATC CTGCAGAAGT CTGTCTTGAG AAGCAGGTAC 600
TCCTGTCACA GCAGAGACAC AGTGGATACC TGAGTAACAA TAATACAAGA CAGGACGTGG 660
GMACAGCAAA AGATTGGGT GTCAGAAGAG GCCGAGAACA CTTYCAGGCA GGAACATTCA 720
RARTTGTCTT TGGAGGAART AGGCMCSAAG GCTGGGCAGG ATTTTCMCGGG GCAGAGATGG 780
AGCAAGCAAT TGAAATGAAA GCCATGGCAT GGGAAAAGGA GCACTGGCCA CAGGGAGTGC 840
AACGTTGTGA TGCAAGGCCA CTGTGGAGCC AT 372

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 312 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

GGCAGAGGCT CACCCAGCA GAGATTGAGG GGAACCGTG ATGAAATTTT TAAGTATCT 60
GCTTGATGAT AATAATTTT CTCTATGTT AATGTTGGCT CCGTTTGGGT GTTTRAGCTTT 120
TGAAAGGAGT ATGAAATGC GGAATGGGGC TTTGGGGCTT GAGGAGGTGT CATCTCTAGT 180
GTTTAAAAAA TTTAATTGCA CAAATAGAAA TAATTCACC ACATTATTGA ACCCCACTAA 240

5 AGCATATCCT TTTTGTCCAT ATTCCCTTTC TGCTGCCCTC GTGTGTACCA TTATTACTCA 300
 GTTGTGATTT GAGCTCGTTC CACTTAAAGT CATTTCATAGA TACTTTTGCG TCGTGTTKGA 360
 ATATTTATMG AATTTCTATT CTGTGTTTTA CTTAATTACT TTATTATGGA ACCTTTACAC 420
 AGGTCTGGTG TACTTGTTCCT TTGAAAAGTC TTATGTTGAC CACCATCACT GAGCATATAG 480
 10 CTTTTCCTTT ATTTCCTTGG GATAATTACC CGAAGTGGAA ATACCGAATC AACTTCTGT 540
 TTTCTTTCTT TGGCACTATT ATATAAATTG TTTTCCAAAC AAGGCATGTT TACAATAGAC 600
 ATTTTTCAAA ATCTGGGTAT TTGTCCTATT TTGCTCTCTG TATGCAGAAT TCAGCGGGGT 660
 15 GCCAAGTCCT TTTCTGTGTG GGTGAGAGA CAGGCTGTGC AGCCCACTGT TGCATAGGAC 720
 TAACTACTAC AAATCATGCT GAGACCGAGC TATTTTTGCT GCTTAGARGC TTTGCAGCCT 780
 20 TGAGTAAGTT TCGNCATCTG GAAACNTTGN AA 812

25 (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1515 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

35 AATTCCGCAC GAGGGAAATT CAAGCACTTT TCCTAAAAGA ACGGGGAATG CATGCTGAAA 60
 CAACACGTTT CCCACAAAGG GAGCAGACAC TGGGCTTGTG AAGCTGCCCC ATACCTTCCC 120
 CACAGAACTG GGGTCCGGCC TCCCTGACAT GCAGATTTCC ACCCAGAAGA CAGAGAAGGA 180
 40 GCCAGTGGTC ATGGAATGGG CTGGGOTCAA AGACTGGGTG CCTGGGAGCT GAGGCAGCCA 240
 CCGTTTCAGC CTGGCCAGCC CTCGCGACCC CGAGGTTGGA CCTACTGTG ACACACCTAC 300
 45 CATGCCGACA CTCTTCAACC TCCTCTGGCT TGCCTGGCC TGCAGCCCTG TTCACACTAC 360
 CCTGTCAAAG TCAGATGCCA AAAAAGCCGC CTCAAAGAGG CTGCTGGAGA AGAGTCAGTT 420
 TTCAGATAAG CCGGTGCAAG ACCGGGGTTT GGTGGTGACG GACCTCAAAG CTGAGAGTGT 480
 50 GGTCTTTGAG CATCCAGCT ACTGCTCGGC AAAGCCCCGG GACAGACACT TTGCTGGGGA 540
 TGTACTGGGC TATGTCACTC CATGGAACAG CCATGGCTAC GATGTCACCA AGTCTTTGG 600
 55 GAGCAAGTTC ACACAGATCT CACCCGTCTG GCTGCAGCTG AAGAGACGTG GCCGTGAGAT 660
 GTTTGAGGTC ACGGCGCTCC ACGACGTGGA CCAAGGGTGG ATGCCAGCTG TCAGGAAGCA 720
 TGCCAAGGGC CTGCACATAG TGCCTCGGCT CCTGTTTGAG GACTGGACTT ACGATGATTT 780
 60

300

CCGGAACGTC TTAGACAGTG AGGATGAGAT AGAGGAGCTG AGCAAGACCG TGCTCCAGGT 840
 GGCAAGAAGC CAGCATTTCG ATGGCTTCGT GGTGGAGGTC TGGAAACGAG TGCTAAGCCA 900
 5 GAAGCGCGTG ACCGACCAGC TGGGCATGTT CACGCACAAG GAGTTTGAGC AGCTGGCCCC 960
 CGTGCTGGAT GGTTCGAGCC TCATGACCTA CGACTACTCT ACAGCGGCATC AGCCTGGCCC 1020
 10 TAATGCACCC CTGTCTGGG TTGGAGCTG CGTCCAGGTC CTGGACCCGA AGTCCAAGTG 1080
 GCGAAGCAAA ATCCTCCTGG GGCTCAACTT CTATGGTATG GACTACGGGA CCTCCAAGGA 1140
 TGGCCGCTGAG CTTGTGTGCG GGGCCAGGTA CATCCAGACA CTGAAGGACC ACAGGCCCCG 1200
 15 GATGGTGTGG GACAGCCAGG YCTCAGAGCA CTTCTTGGAG TACAAGAAGA GCGCCAGTGG 1260
 GAGGCACGTC GTCTTCTACC CAACCTGAA GTCCCTGCAG GTGCGGCTGG AGCTGGCCCC 1320
 GGAAGCTGGG GTTGGGCTCT CTATCTGGGA GCTGGGCCAG GGCCTGGACT ACTTCTACGA 1380
 20 CCGCTCTAG GTGGGCATTG CCGCCTCCGC GGTGGACGTC TTCTTTTCTA AGCCATGGAG 1440
 TGAGTGAGCA GGTGTGAAAT ACAGGCCTTC ACTCCGTTAA AAAAAAAAAA AAAAAAAAAA 1500
 25 AAAAAAAAAA AAAAA 1515

30 (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 704 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

40 AAGATGGTGG CGCCCAGAGC TTGGCTCTAT GCTGCTCCCC TGAGAGAGGC GTTTCCTCA 60
 ACCAGTTTTG CAAGGAGTTC AATGAGAGGA CAAAGGACAT CAAGGAAGGC ATTCTCTGCG 120
 CTACCAAGAT TTTAGTGAAG CCTGACAGGA CATTTGAAAT TAAGATTGGA CAGCCCACTG 180
 45 TTTCCTACTT CCTGAAGGCA GCAGCTGGGA TTGAAAAGCG GGGCCGGCRA ACAGGGAAG 240
 AGGTGGCAGG CCTGGTGACC TTGAAGCATG TGTATGAGAT TGCCCGCATC AAAGCTCAGG 300
 50 ATGAGGCATT TGCCCTGCAG GATGTACCCC TGTGCTCTGT TGTCCGCTCC ATCATCGGGT 360
 CTGCCCCGTC TCTGGGCATT CCGGTGGTGA AGGACCTCAG TTCAGAAGAG CTTCAGCTT 420
 TCCAGAAGGA ACGAGCCATC TTCTGGCTG CTCAGAAGGA GGCAGATTG GCTGCCCCAAG 480
 55 AAGAAGCTGC CAAGAAGTGA CCCTTGCCCC ACCAACTCCC AGATTTCAA GGAGGTAGTT 540
 GCAAAAGCTG TGCCCAAGGG GAGGAAGSAG GTCACACCAA TATGATGATG GTTTTCATGA 600
 60 CTTTGAATGA TATATTTTGG TACATCTAGC TGTATCGAGG CATCAGGCTT GAATTAACAT 660

CCTTTCTTAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA

704

5

(2) INFORMATION FOR SEQ ID NO: 41:

10

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1094 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGCAGCTTTC TTACAAACCC ATCCTTCTGA AATGTTGCTT CAAATTCATC CTCTGCTCCC	60
CAGTCCCACT ATTCCACACA TACTGTTACT GTTTCCTTAT CCTACTTTCCT CAATTTTGGG	120
ACATAGTTGC AGTTACTGCA TTGAATACCT GTGGGTTTGC CTGTTGTTCT GTCTGTCTCT	180
GTGGTTCTTG TAATANTGGA TCCCAGAGAT AAAATGGACA GTTGTNATGC ACAGTTAATT	240
CAGAAACTAG ACCTTACTTG CTGTGTGAAA TACCAACTAA ATTCTCAGTG AACTCAGCTG	300
ANCTTTATCT CCTTTTGTTC CCCCATTATA TAATTTTCACT TCAGGCCCCAG AAAGATGGAA	360
TCCCAGCTAA GAAATACAAG TTACACCCCTG TACTAGCAGC CCATGTGTGC ATGTTCTTTA	420
AGTGCTCTTG CAGCTATGTC ATTTATATTG ATTTCCCTGT ATTATTATAA GCAAAGCAA	480
TTTGAGGAAA AAAACCCATA ATACCACACC TCATTTTTTT CAAGTAATAG GGTGATAAGT	540
CTCATCTTTC ATATAATATG TTGAGTATGC AGTATATTAT GTGTTAGGCT CTGGANAGGC	600
AGAGGTTAGA TCATGTWACA GATCATATCK GATTAGGCAG ATAAACAGTA TTTTAACCTT	660
TTCCCTATTA TATGTAACCT GCTTTCAGGT TTTTAAATGT TACTATTATG TCTTTAATAT	720
ATTATCTTTA TTTGTACTTT TGTATACAGA GTGATTTTCC TTTTAAAAA AAAATTGTGT	780
CTTTAGGATG GATTCCAAAG ATGTGGAATC AGTAGGTTTA AGGAATATGG ATATTTTGGC	840
TGGCAAGGTG GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAGGTG GGTGGATCAC	900
CTGAAGTCAG GAGTTCGAGA CCAGCCTGAC CAACATGGCG AAACCCCTGTT TTTACTAAAG	960
ACACACWAA AATTGCCAG TGGTGGTGGC ATGTGCTTGT AGTCCCACTT AGCTACTCGA	1020
GAGGCTGAGG CAGGAGAATC GCTTGAACCC GGGAGGCAGA GGTTCAGTG AGGCAAGATG	1080
GCACCTCTAC ACTC	1094

55

(2) INFORMATION FOR SEQ ID NO: 43:

60

(1) SEQUENCE CHARACTERISTICS:

302

(A) LENGTH: 1821 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(E) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	TGGCTTAGGC GAGCAGCCTT CCGTTGGCTG GAAGTACTGG ACAGACCCCTT TTGAGATGTG	60
10	CCTGTGGTGC TTGCGAGATG TGTGTAGTGG TCTTAGCTCT TTGTTGAGCT TGTGTGTGTG	120
	TTGTGTATTC TTAGCTGTAT GGTGAATTTG GCGGTGTGTT GGAGGGCTTC TTAGCTCTTT	180
15	GCTGAGATTC TATTCTATG TGTTTGTATC ASCTGAATGT TGCTGGAAAT AAAACCTTGG	240
	TTTGTGAGG CTGTTTTTTC TGGGAAGTAA GTACGGGAAA AGGTCTTTGA GGGTTCCTAG	300
	GCTCCTTTGT ACAACGGGAA AATGCTCAA AGCCTTGCTT CCCAGCAACT TGGGGCTGGT	360
20	TCCAGTCCC TGGTCTGCC CTTGCTGGT TCTTATCTCA AGGCAGAGCT TCTGAATTTT	420
	AGGCCTTCAT TCCAGGCCCC TCTTGTGGCC AGGCCTTCCT TTGCTGGAGG AAGGTACACA	480
25	GGGTGAAGCT GATGCTGTAC TTGGGGGATC TCTTTGGCCT GTTCCACCAA GTGAGAGAAG	540
	GTAATTACTC TTGTACCTCC TGTTCAGCCA GGTGCATTAA CAGACCTCCC TACAGCTGTA	600
	GAATTAATTC TCCAGAGCT GAGGCAAGGG GATTCTCTAG CTCATTTCGA GAACAAGTGC	660
30	TTTATATATA GTTTAAAGTA GTAATGCTA CTGTATTTAG TGGGGTGGAA TTCAGAAGAA	720
	ATTTGAGGAC CAGATCATGG GTGGTCTGCA TGTGAATGAA CAGGAATGAG CCGGACAGCC	780
35	TGGCTGTGAT TGGTTCTTC CTCCCGATT CTGACCTTCT CTGCCCCTTAC ATTTTGTGTT	840
	CTCCATCTAC CAGCATCCAC CAGTCTATTT ATTAAGTTAG CAAGAGGACA AGTAAAGGGC	900
	CGCTTTGGCT TGAATTTGCT TCTTTCTTTC TGTGGAGGAT ATACTAAGTG CGACTTTGCC	960
40	CTACCTTATT TGGAAATCCC TAACGAATT GAGTTTCTA TTAAGGATCC AAAAAGAAAA	1020
	ACAAAATGCT AATGAGGCA TCAGTCAAGG GTCACATGCC AATAAACAT AAATTTTCCA	1080
45	GAAGAAATTA AATCCACTA GACAAATAAA GTAGAGCTTA TGAAATGGTT CAGTAAGGAT	1140
	GAGTTTGTG TTTTGTGTT TGTTTGTTT TGTTTTTTTA AAGACGGAGT CTCGCTCTGT	1200
	CACTCAGGCT GGAATGCAAT GGTATGATCT TGGCTCACTG TAACCTCCGC CTCGCCGGTT	1260
50	CAAGCCATTC TCTGGCTCA GTCTCTGAG TAGCTGGGAT TACAGGTGCG TGCCACCATG	1320
	CCTGGCTAAT TTTGTGTTT TTAGTAGAGA CAGGGTTTCA CCATGTTGGT CGGGCTGGTC	1380
55	TCAAACTGCT GACCTCTTGA TCCGCTGCCC TTGGCTCCC AAAGTGATGG GATTACAGAT	1440
	GTGAGCCACC CCGGCTTAC CCAAGGATGA GATTTTAA -GTATGTTTCA GTTCTGTGTC	1500
	ATGGTTGGA GACAGAGTAG GAAGGATATG GAAAAGGTCA TGGGGAAGCA GAGTGATTC	1560
60	ATGGCTCTGT GAATTTGAGG TGAATGTTT CTTATGTTCT AGGCCACTTG TGAAGAATAT	1620

GAGTCAGTTA TTGCCAGCCT TCGAATTAC TTCTCTAGCT TACAATGGAC CTMTTGAAC 1680
 GGAAAACACC TTGTCTGCAT TCACCTTAAA ATGTCAAAC TAATTTTAT AATAATGTT 1740
 TATTTTCACA TTGAAAAAAA AAAAAATTT AAAAACYCGG GGGGGGCCCC G#ACCCCAT 1800
 NGCCCCCTAAG GGGGGGGGT T 1821

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1024 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGCGACAGT TGAAGAAGCG ACCGAGGGAC TCGGAGTCGT TACTGAGGAT GACCGGGCAT 60
 GGCAAGAAGT GCACCGCAGG GCGGTCTACA CCTACCACGA GAAGAAGAAG GACACAGCGG 120
 CCTCGGGCTA TGGGACCCAG AACATTCGAC TGAGCCGGGA TGCCGTGAAG GACTTCGACT 180
 GCTGTTGTCT CTCCCTGCAG CCTTGCCACG ATCCTGTTGT CACCCAGAT GGCTACCTGT 240
 ATGAGCGTGA GGCATCTG GAGTACATTC TGCACCAGAA GAAGGAGATT GCCCGGCAGA 300
 TGAAGGCCTA CGAGAAGCAG CCGGGCACCC GCGCGGAGGA GCAGAAGGAG CTTCAGCGGG 360
 CGGCCTGGCA GGACCATGTG CCGGGCTTCC TGGAGAAGGA GTCGGCTATC GTGAGCCGGC 420
 CCCTCAACCC TTTCACAGCC AAGCCCTCT CCGGCACCAG CCCAGATGAT GTCCACCTG 480
 GGCCCACTGT GGTCTCTCCA AGTAAGGACA AGGACAAAGT GCTGCCACG TTCTGGATCC 540
 CGTCGCTGAC GCCCGAAGCC AAGGCCACCA AGCTGGAGAA GCGCTCCCGC ACGGTGACCT 600
 GCCCCATGTC AGGGAAGCCC CTGCCATGT CCGACCTGAC GCGCGTGAC TTCACACCGC 660
 TAGACAGCTC CGTGGACCGC GTGGGGCTCA TCACCCGAG CGAGCGCTAC GTGTGTCCG 720
 TGACCCCGGA CAGCCTGAGC AACGCCACCC CCGCGCTGT GCTGCGGCCC TCTGGGGCTG 780
 TGGTCACCTT CGAATGCGTG GAGAAGCTGA TTCGGAAGGA CATGGTGGAC CCTGTGACTG 840
 GAGACAACT CACAGACCGC GACATCATCG TGCTGCAGCG GGGCGGTACC GSTTCGCGGG 900
 CTCCGGAGTG AAGCTGCAAG CGGAGAAATC ACGCCCGGTG ATGCAGGCCT GAGTGTGTGC 960
 GGGAGACCAA ATAAACCCGC TTGGGTGCGG AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1020
 AAAA 1024

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 983 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

CGACACGGCT GCGAGAAGAC GACAGAAGGG CCGGACCGCG AGCCGTCCAG GTCTCACTGC 60
 TGTGCCCCCC CCAGAGCCTA GAGGATGTTT CATGGGATCC CAGCCACGCC GGGCATAGGA 120
 GCCCCCTGGGA ACAAGCCCGA GGTGTATGAG GAAGTGAAGT TGTACAAGAA CGCCCGGGAG 180
 AGGGAGAAGT ACGACAACAT GGCAGAGCTG TTTGCGGTGG TGAAGACAAT GCAAGCCCTG 240
 GAGAAGGCCT ACATCAAGGA CTGTGTCTCC CCCAGCGAGT AACTGCGAGC CTGCTCCCGG 300
 CTCTGGTCC AATACAAAGC TGCTTCAGG CAGGTCCAGG GCTCAGAAAT CAGCTCTATT 360
 GACGAATTCT GCGCAAGTT CCGCCTGGAC TGCCCGCTGG CCATGGAGCG GATCAAGGAG 420
 GACCGGCCCCA TCACCATCAA GGACGACAAG GGCAACCTCA ACCGCTGCAT CGCAGACGTG 480
 GTCTCGCTCT TCATCACGGT CATGCACAAG CTGCGCCTGG AGATCCGCGC CATGGATGAG 540
 ATCCAGCCCCG ACCTGCGAGA GCTGATGGAG ACCATGCGCC GCATGAGCCA CCTCCACCCC 600
 GACTTTGAGG GCGCCAGAC GGTGAGCCAG TGGCTGCAGA CCTGAGCGG CATGTCCGGG 660
 TCAGATGAGC TGGACGACTC ACAGGTGCGT CAGATGCTGT TCGACCTGGA GTCAGCCTAC 720
 AACGCCTTCA ACCGCTTCCT GCATGCCTGA GCGCGGGCCA CTAGCCCTTG CACAGAAGGG 780
 CAGAGTCTGA GCGATGGCT CCTGGTCCCC TGTCCGCCAC ACAGGCCGTG GTCATCCACA 840
 CAACTCACTG TGTGCAAGT CCTGTCTGGT GTCTGTCTTT GGTGTGAGAA CTTTGGGGCC 900
 GGGCCCTCC CCACATAAA GATGCTCTCC GACCTTCAA AAAAAAAAAA AAAAAAAGR 960
 KSGGGCGGGT CCCCANTCCC CCC 983

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2421 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

CCGGCTGATC GCTGCGCGTC CGCCAATACA ATAGAGCCAK CCACTACCAG CAGCCTGGCC 60

	CTCTTCTCTCC TTCTCCAGAG AGACCAATCC AGCCGAAGTC GGGCTTTGCC TGAGGAGAAG	120
	GAGGAAGTGA CCATGGACAC AAGTGAAAAC AGACCTGAAA ATGATGTTCC AGAACCTCCC	180
5	ATGCCATTATG CAGACCAAGT CAGCAATGAT GACCGCCCGG AGGGCAGTGT TGAAGATGAG	240
	GAGAAGAAAG AGAGCTCGCT GCGCAATCA TTCAAGAGGA AGATCTCCGT TGTCTCAGCT	300
10	ACCAAGGGGG TGCCAGCTGG AAACAGTGAC ACAGAGGGGG GCCAGCCTGG TCGGAAACGA	360
	CGCTGGGGAG CCAGCACAGC CACCACACAG AAGAACTTT CCATCAGTAT CACCACTGAA	420
	TCACTAAAGA GCTTCATCCC CGACATCAAA CCGCTGGCGG GGCAGGAGGC TGTGTGGAT	480
15	CTTCATGCTG ATGACTCTCG CATCTCTGAG GATGAGACAG AGCGTAATGG CGATGATGGG	540
	ACCCATGACA AGGGGCTGAA AATATGCCCG ACAGTCACTC AGGTAGTACC TGCAGAGCGC	600
20	CAGGAGAATG GGCAGAGGGA AGAAGAGGAA GAAGAGAAGG AACCTGAAGC AGAACCTCCT	660
	GTACCTCCCC AGTGTCTAGT AGAGGTGGCC TTGCCCCCAC CTGCAGAGCA TGAAGTAAAG	720
	AAAGTGACTT TAGGAGATAC CTTAACTCGA CGTTCCATTA GCCAGCAGAA GTCCGGAGTT	780
25	TCCATTACCA TTGATGACCC AGTCCGAAGT GCGCAGGTGC CCTCCCCACC CCGGGGCAAG	840
	ATTAGCAACA TTGTCCATAT CTCCAATTG GTCCGTCTTT TCACTTTAGG CCAGCTAAAG	900
30	GAGTTGTTGG GCGGCACAGG AACCTTGGTG GAAGAGGCCT TCTGGATTGA CAGATCAAA	960
	TCTCATTTGCT TTGTAACGTA CTCAACAGTA GAGGAAGCTG TTGCCACCCG CACAGCTCTG	1020
	CACGGGGTCA AATGGCCCCA GTCCAATCCC AAATTCCTTT GTGCTGACTA TCGCGAGCAA	1080
35	GATGAGCTGG ATTATCACCG AGGCCTCTTG GTGGACCGTC CCTCTGAAAC TAAGACAGAG	1140
	GAGCAGGGAA TACCACGGCC CCTGCACCCC CCACCCCGAC CCGCGGTCCA GGCACCACAG	1200
40	CACCCCCGGG CAGAGCAGCG GGAGCAGGAA CGGGCAGTGC GGAACAGTG GGCAGAACGG	1260
	GAACGGGAAA TGCAGCGGCG GGAGCGGACT CGATCAGACC GTGAATGGGA TCGGACAAA	1320
	GTTCGAGAAG GCGCCCGTTC CCGATCAAGG TCCCGTACC GCGCCGCAA GGAACGTGCG	1380
45	AAGTCTAAG AAAAGAAGAG TGAGAAGAAA GAGAAAGCCC AGGAGGAACC ACCTGCCAAG	1440
	CTGCTGGATG ACCTTTTCCG AAAGACCAAG GCAGTCCCT GCATCTATTG GCTCCCACTG	1500
50	ACTGACAGCC AGATCGTTCA GAAAGAGGCA GAGCGGCGG AACGGGCGCA GGAGCGGGAG	1560
	AAGCGCGAA AGGAGCAAGA AGAAGAAGAG CAAAGGAGC GGGAGAAGCA AGCCGAGCGG	1620
	GAACCGAACC GACAGCTGGA GCGAGAGAAA CGTCGGGAGC ACAGTCGGGA GAGGGACAGG	1680
55	GAGAGAGAGA GAGAAAGGA GCGGGACAGG GGGGACCGAG ATCGGGATAG GGAAAGGAC	1740
	CGAGAACCAG GCAGGGAAAG GGATCGCAGG GACACCAAGC GGCACAGCAG AAGCTGGAGT	1800
60	CGGAGCACAC CTGTCCGGGA CCGGGGTGGG CGCCGCTAGC TGGGAAACA CTAGAGCTGC	1860

306

AGGTACCAGC CACTCGGCCC CAGGGGGTTA TGGCCACAGA GGGATAGGCA CAGTCTCCAC 1320
 CACCTTGGAG CCAAGGGTCT TTCACATCAC CTATCCCTAC ATACATACCA AATGGAAAG 1330
 5 TGGCCATCCT TTTCCCCCA AACACACCCC CTTAACCTAT CTCTTGGGAC TTAGCCCGAC 1340
 CCTCCCTCTC ATTTCCCAT T AAGTCTGAGA GGCAAGAGCT AGGTTAGGCA AGGAGGTGCT 2100
 TGGCCACAGA TGGGGAACAG CCAGGTGCCC CAGTCTCTG ATTTTCTCTC CATCCTGCTT 2160
 10 ACCACCTCCC TGGGTACTTA CAGCCTTCTC TTGGGAACAG CCGGGGCCAG GACTGGGTCA 2220
 CCTATGAGCT GAATCAGCAT CTCCTCTGA GTCCGAGGCC CCTGCACTT CCCAGTCTCT 2280
 15 TCTGTCTGC AGCCCTTGGC TCTTCCAC AGGTTCCACT TTATATCCAC CTTTCTCTT 2340
 TGTTCAATTT TTATTTTAT TTTTATTATT ATTAAATGAT GTGTCTATG GAAAAAATA 2400
 TAAAAATCTG ACTTAGTTTT A 2421
 20

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 840 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CTCAAACTCC TGAGCTGAAG CGATCTACCT GCCTCAGCTA GCATTACAGG TGTGAGCCAC 50
 35 CGCACCCCAAC CTCAATAAGC KTATTTGATA AAKATATGC AAGCTCCCTT TATKCACTTT 120
 TCATTCAGAA TGTTTAGTAA TTTGTATGT TTTTCAGATT TTCAGCCCA TATATCTCTT 180
 40 TGCCCACTGT GTCACTGTAT TCTACCTAWA CATCATCAGG TGTTCCTGCT ATTGGCTGTA 240
 TGATGGAACA CTGCGGCTCA TTTCTTGAA AACTGCCGAT AGTGCATAGA RTGCTGGGAT 300
 GGAAACCAGA ARCTTTGAAT TCAAGCCTTG GTTCTGCCTT GTTTTTCCTT GGGTGGCCTT 360
 45 GAGTCAGCCA CATACCTTTT AAAATCTCAA TTTATTAGAA ATTATTCCAA ATCAAAATCA 420
 AATGAGAAGG TATATACAAA AGTGCTTTAT CCCACAATAA ACTATTCAAG AGAGAGCCAA 480
 50 GGAGAGGACA TTTACTCAAC ACCTCCTAAA AGGCAGCCAG TGAAATTAGG CATTTTATTT 540
 AATCCTCCTG GCAACTCTGA GAGTAAGCA TTATTAATCC CATTTTGGCT GTTTAAAGCA 600
 ATTATTGCA CTAGATTCCA GTGTAGTTT AGYTTGAGAA AAAAAATCC TGAGATGTGA 660
 55 ATTACAGCT TTCTGGGTTT AAAGCCCAAG CTCTATCACA TCATGCTATT ATTGTTACAT 720
 TACTGCTAGT TCTATGAAAA GAAATACTAA TTTATGAAAT ACATCTTATC CAAAAAATA 780
 60 AAAAAAATC TGGGAGGGGG GGGCGTACC CAAATCGCCG GATAGTGAT GTAAACATC 840

5 (2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2432 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

15 GCCACGAGGC CCGAACGCT GAGGAAGGTC CGTCCCGCC TTCCCGGGCG CGCCATGGAG 60
CCCCGGGGCG TTGCAGAAGC CGTGGAGACG GGTGAGGAGG ATGTGATTAT GGAAGCTCTG 120
CGGTATACA ACCAGGAGCA CTCCAGAGC TTCACGTTTG ATGATGCCCC ACAGGAGGAC 180
20 CGGAAGAGAC TGGCGGASTG CTGGTCTCCG TCCTGGAACA GGGCTTGCCA CCTCCCAAC 240
GTGTCACTCG GCTGCAGAGT GTCCGAATCC TGTCCCGGGA CGCRACTGC CTGGACCCGT 300
25 TCACCAGCGG CCAGAGCCCTG CAGGCAYTAG CCTGYTATGY TGACATCTCT GTCTCTGAGG 360
GGTCCGTCCT AGACTCCGCA GACATGGATG TTGTACTGGA CTCCTCAAG TGCCTGTGCA 420
ACCTCGTGCT CAGCAGCCCT GTGGCACAGA TGCTGGCAGC AGAGGCCCCG CTAGTGGTGA 480
30 AGCTCACAGA GCGTGTGGGG CTGTACCGTG AGAGGAGCTT CCCCCACGAT GTCCAGTTCT 540
TTGACTTGCG GCTCCTCTTC CTGTAACGG CACTCCGCAC CGATGTGCGC CANAGCTGTT 600
35 TCAGGAGCTG AAAGGAGTGC GCCTGCTAAC TGACACACTG GAGCTGACGC TGGGGGTGAC 660
TCCTGAAGGG AACCCCCCAG CCACGCTCCT TCCTTCCCAA GAGACTGAGC GGGCCATGGA 720
GATCCTCAAA GTGCTCTTCA ACATCACCCT GGACTCCATC AAGGGGGAGG TGGACGAGGA 780
40 AGACGCTGCC CTTTACCGAC ACCTGGGGAC CTTTCTCCG CACTGTGTGA TGATCGCTAC 840
TGCTGGAGAC CGCACAGAGG AGTTCCACGG CCACGCAGTA ASCCTCTGG GGAACCTGCC 900
45 CCTCAAGTGT CTGGATGTTG TCCTCACCTT GGAGCCACAT CGAGACTCCA CGGAGTTGAT 960
GGGAGTGAAT ATGGATGTGA TTCGTGCCCT CCTCATCTTC CTAGAGAAGC GTTTCACAA 1020
GACACACAGG CTGAAGGAGA GTGTAGCTCC CGTCTGAGC GTGCTGACTG AATGTGCCCG 1080
50 GATGCACCGC CCAGCCAGGA AGTTCTTGAA GGGCCAGGTG CTGCCCCCTC TGGGGGATGT 1140
GAGGACACGG CCTGAGGTTG GGGAGATGCT CCGGAACAAG CTGTCCCGCC TCATGACACA 1200
55 CCTGGACACA GATGTGAAGA GGGTGGCTGC CGAGTTCTTG TTTGTCTGT GCTCTGAGAG 1260
TGTGCCCCGA TTCATCAAGT ACACAGGCTA TGGGAATGCT GCTGGCCTTC TGGCTGCCAG 1320
GGGCTCTATG GCAGGAGGCG GGGGAGGCG AGTACTCAGA GGATGAGGAC ACACACACAG 1380

5 ATGAGTACAA GGAAGCCAAA GCCAGCATAA ACCCTGTGAC CCGGAGGGTG GAGGAGAAGC 1440
 CGCCTAACCC TATGGAGGGC ATGACAGAGG AGCAGAAGGA GCACGAGGCC ATGAAGCTGG 1500
 10 TGACCATGTT TGACAAGCTC TCCAGGAACA GAGTCATCCA GCCAATGGGG ATGAGTCCCC 1560
 GGGGTCATCT TACGTCCCTG CAGGATGCCA TGTGGGAGAC TATGGAGCAG CAGCTCTCCT 1620
 CCGACCCCTGA CTCGGACCCCT GACTGAGGAT GGCAGCTCTT CTGCTCCCCC ATCAGGACTG 1680
 GTGCTGCTTC CAGAGACTTC CTTGGGGTTG CAACCTGGGG AAGCCACATC CCACTGGATC 1740
 CACACCCGCC CCCACTTCTC CATCTTAGAA ACCCCTTCTC TTGACTCCCG TTCTGTTCAT 1800
 15 GATTTGCCTC TGGTCCAGTT TCTCATCTCT GGA CTGCAAC GGTCTTCTTG TGCTAGAACT 1860
 CAGGCTCAGC CTCGAATCC ACAGACGAAG TACTTTCTTT TGTCTGGGCC AAGAGGAATG 1920
 TGTTCAGAAG CTGCTGCCCTG AGGGCAGGGC CTACCTGGCC ACACAGAAGA GCATATGGGA 1980
 20 GGGCAGGGGT TTGGGTGTGG GTGCACACAA AGCAAGCACC ATCTGGGATT GGCACACTGG 2040
 CAGAGCMANT GTKTTGGGGT ATGTGCTGCA CTTCCCAGGG AGAAAACCTG TCAGAACTTT 2100
 25 CCATACGAGT ATATCAGAAC ACACCCCTCC AAGGTATGTA TGCTCTGTTG TTCCTGTCTC 2160
 GTCTTCACTG AGCCGAGGGC TGGAGGCCTC TTAGACATTC TCCTTGGTCC TCGTTGAGCT 2220
 GCCCACTGTA GTATCCACAG TGCCCGAGTT CTCGCTGGTT TTGCCAATTA AACCTCCTTC 2280
 30 CTACTGGTTT AGACTACACT TACAACAAGG AAAATGCCCC TCCTGTGACC ATAGATTGAG 2340
 ATTTATACCA CATACCACAC ATAGCCACAG AAACATCATC TTGAAATAAA GAAGAGTTTT 2400
 35 GGACAAAAAA AAAAAAAAAA AAAAAAAAAA AA 2432

40 (2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1742 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

50 GTCTGCAGG AGCTGCACGC GCGCGAGGTG CGCANGAACA AGGAGCAGCG AGAAGAGATG 60
 TCGGGCTAAG GCGCCGGSAC GCGSGGCGCC CATCCTGCCA CGGAACACGT TCGGGTTTTG 120
 55 GTTTTGTTC GTTCACCTCT GTCTAGATGC AACTTTTGTT CCTCCTCCCC CACCCAGCC 180
 CCCAGCTTCA TGCTTCTCTT CCGCACTCAG CCGCCCTGCC CTGTCTCTGT GSTGAGTCGC 240
 TGACCAGGCC TTCCCTTGCA GGAGCCGCGG GCGGTGRAGA CCGGTTCCTT CGGTGCAGAC 300
 60 ACCAGGCGCG GCGCGCTGG GTCCCCCGGG GGCCTGTGA GAGAGGTGGT CGTGACCGTG 360

	GTAAACCCAG GCGCGTGGCG TGGGATCRCG GGTCTTTACG CTGGGCTGTC TGGTCAGCAC	420
5	GTGCAGGTCA GGGCAGSTCC TCTGAGCCCG CCCCCCTGGC CAGCGGGCGA GGCTACAGTA	480
	CCTGCTGTCT TTCCAGGGGG AAGGGGCTCC CCATGAGGRA GGGGCGACCG GCGACGGGGG	540
	TGATGGTGCC TGGGAAGCCT GCKTGTGCAN CCGGTGCTTG TTGAACTGGC AGGCGGGTGG	600
10	GTGGGGGCTG CAGCTTTTCT TAATGTGGTT GCACAGGGGT CTTCTRAGAC CACCTGGCGT	660
	GAGGTGACAC CCCTGGGCCT TCCTGGAAGC CTGCAGTTGG GGGCCTGGCC TGAGTCTGCT	720
15	GGGAGTGGG CATTTCTCTGC CAGGGACCCA TGACCAGGCT GCATGGTCTA GAGTTGTGG	780
	GCAGCATGGA CAGTCCCCCA CTCAGAAGTG CAAGAGTTCC AAAGAGCCTC TGGCCCAGGC	840
	CCCTCGGTGG GACAGCCCCG CCGCCCTTCC CCACCAGGCC TTTGCAGATG TCCTTGAAAG	900
20	ACCCACCCCTA GAGCCCTTTG GAGTGTGGC CCCTCCTGTG CCCTCTGCCC TGGTGAAGC	960
	GGCASCACAA GTCTCTCTCA GGGAGCCCCA AGGGGGATT TKTGGGACCG CTGCCCCACG	1020
25	ATCCAGGTGT TGAAGGGCA GCGGGTAAGG TTCCCAAGCC AGCCCCAACA CCCTTCCCAC	1080
	TTGGCACCCA GAGGGGGCTG TGGGTGGAGG CCTGACTCCA GGCCTCTCCT GCCCACACCC	1140
	TCTGGGCTGA GTTCTTTCTT TCCCTTGGAC GCCCAGTGCT GGCCTTGGAG GACGGTCAGC	1200
30	TGGAGGATGG CCGTGGGGA GGCTGTCTTT GTACCACTGC AGCATCCCCC ACTTCTCCAC	1260
	GGAAGCCCCA TCCCAAAGCT GCTGCCCTGGC CCCTTGCTGT AAAGTGTGAA GGGGGCGGCT	1320
35	GAGTCTCTTT AGGACCCAGA GCCAGGGCCC TCAACTTCCA TCCTGCGGGA GGCCTTGGCC	1380
	GGGCACTGCC AGTGTCTTCC AGAGCCACAC CCAGGGACCA GCGGAGGATC CTGACCCCTG	1440
	CAGGGCTCAG GGGTCAGCAG GGACCCACTG CCCCATCTCC CTCTCCCCAC CAAGACAGCC	1500
40	CCAGAAGGAG CAGCCAGCTG GGATGGGAAC CCAAGGCTGT CCACATCTCG CTTTGTGGG	1560
	ACTCAGAAAG GGAAGCAGAA CTGAGGGCTG GGATATTCTT CATGGTGGCA GCGTTCATAG	1620
45	CGAAAGCCTA CTGTAAATATG CACCCATCTC ATCCACGTAG TAAAGTGAAC TTAATAATTG	1680
	AATCAAATGA ACATTAAT AAACACCTGT GTTTTAAAGA AAAAAAAAAA AAAAAAACTG	1740
50	CG	1742

(2) INFORMATION FOR SEQ ID NO: 50:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1487 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

GGCACGAGCC TCCGCGAAGT GTGGAGTCGG CGGAGGGCTG GAATCAGCGT GGGCTCCAGC 60
 5 TCGCTGGCAG CCGGGTGGCA GAACTCTTCC GAGGCTCCTT GGAAGAAGC TACACCCGAG 120
 GGAGCCCGAT GGGCTCGAA AACCTGGCCC GCTCTGGTTC TGTACCATG CAAGGGGAAC 180
 10 CGTAAACTGA GCTTTTCTAA CGTGGGTTTC TGCCAAGTAC TTTCCAGCT GCGCCCTTCC 240
 CCCCAGCACA CAGGAGAGCC TCTGTGTAGC CAGCGCTTGA CAGTCGTTAG GTAGGTTGTA 300
 CTGTGTAGGG AGGAGCTCAA GATCATGAAT GGTGTGCACA GGAGAAAGCG GTTGCATCTT 360
 15 TGCAAAACTA TATACCTGCT GTGTTTGTG TTTTCTTTTC TGCTGAGTAA TGAAGTTGTA 420
 AGTTCACACT GGCACATTCT CAGGGCTGTG CAGATTATTT GCACTTTATT TCATAGGTG 480
 20 ATAAGTGCTT TTTAGCTTTC TTTGTATATT GAGTTGCTTT TGAATTGCTT CCCATATTTT 540
 TATTTCTATC AACTGAACA ATTGTGGCCC CTCTATTTTA TTTATAAAGG TTCAGTGTAT 600
 CTTTGCCTGC CTACATCAAT CTGCAAGGGA GTTGCAGAAA GCCTCATGTT CATCGAGCGG 660
 25 TGAGTCACAA CCAATTCTTA AGCTGTTATA ACAAAAAAGT GTTTGCTTTT TTTCAAGA 720
 AACTTTAAAA GTGTAGTTTA GAAAGAAAAC ATTTTCAATA AAAAGACACT ACATTAATCC 780
 30 TGGATGCTTG CAAATCCTAA AATMTATTCC TCCTCTAGCG TTGCACAGCT CTGTGTTGTA 840
 TACACAGACT AGCTTTAAAA TTTGTCACAT ACCACTTTAC CTTTACTTTT ATGTATCATT 900
 CCCCCGACTT CTTACTGCA GGTGTGGGCA AGAAAACCTT TCCTTTAACA CTTTTCACA 960
 35 CCGGGCATAA AATTCTGCAG CTGAGGTCTT GAAGAATGCA GATGGGTACA GTATGTGTTG 1020
 GAGCTCACAG TGTGTATTGA CTAACCTAGT TCCTTTTNTG CTTTNTTGG TATGTCTTG 1080
 40 TTAAAAGTGA CTCCCAGGTA GCAACTCTCT TTTTAAAGG TGGGAACGAA AGGGACGTAG 1140
 GAAGAATAGA TCTAGATTAT TTAACAGTCT TCGATAGAGT TTGAAAGCTT TCTTCTTCAT 1200
 TCAATTTTGG GCAAAATACT GCCTCTGCAT TTGTTTATAA CAAAAGATT AGATTAATAA 1260
 45 GTAGCTTTTG TTGGTGAAA TTACCAGCTC TATAAGTCAC CCTTGGTGGT TCATGGACCT 1320
 CTGATTAGCT TGGGTTTTC AGTCTCATTC CCACATGTAT ATGTGGAGCC AATGGCCTTT 1380
 50 TGGTGCTCAG CTGTTTACGT CTGACTCCTT GACTTCTTTG GTACAGTGAT GGAGTCAGAT 1440
 CTCATTAAGT GTGATTCTCC ATGGATATAA CCAGCCCCAA AAAAANG 1487

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1328 base pairs

(B) TYPE: nucleic acid

311

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

5 GGCACGAGCT CGTGCCGAAT TCGGCACGAG AGAAGATTTG AAGAAGCCAG ATCCAGCTTC 60

CCTGCGGGCT GCTTCTTG TG GGGAGGGAA AAAGAGGAAG GCCTGTAAGA ACTGCACCTG 120

10 TGGCCTTGCC GAAGAAGTGG AAAAAGAGAA GTCAAGGGAA CAGATGAGCT CCCAACCCTA 180

GTCAGCTTGT GGAAACTGCT ACCTGGGCGA TGCCTTCGGC TGTGCCAGCT GCGCCTACCT 240

15 TGGGATGCCA GCCTTCAAAC CTGGGGAAAA GGTGCTTCTG AGTGATAGCA ATCTTCATGA 300

TGCTTAGGAG GTTCTGACA TGGGACCCAT CTGCTCTCC AGCCAACTCC TGTCCCTCAC 360

ATCCACCCT GGTGGCTCTT CCCACCTCTT CTGGATTGT TCACTCTGAG ATCTGTTTGC 420

20 AGAGTGGGTG CTTAGCAGAC AGAGTGAAGT TGGCTGGGG GCACAGTGGT GTGTAGTGCT 480

GCTGTGTATC AAAAGACCAA GGTATTATGG GACCTGTTT CAGAATGGGA TGGGTTTCTT 540

CACCTCATGT TAAGAGAAGG GAGTGTGTCC TGAAGAAGCC CTCTCTCTGA TGTAAATG 600

25 CTGACCAGAA CGCTCTTGAG CCCAGGCATC GTTGAGCATT AACACTCTGT GACAGAGCTG 660

CAGACCCCTG CCTTGAGTCT CATCTCAGCA ATGCTGCCAC CTTCTTGTCT TTCAGAGTTG 720

30 TTAGTTTACT CCATTCTTTG TGACACGAGT CAAGTGGCTC ACAACCTCT CAGGGCACCA 780

GAGGACTCAC TCACTGGTTG CTGTGATGAT ATCCAGTGTG CTTCTGCCCC CTTCATCCC 840

35 CAACCACATT TGAAGTAGC ATTGCATCTG TGTCTGTG TGATTTATGT TAACCTTCAG 900

GTATTAAACT TGCTGCATAT CTTGACATAT CTTGAGATTG TGCATGTCTT GTAAAGAGAG 960

GGGATGTGCA TTTGTGTGTG ATGTTGGATA GTCATCCACG CTCAGTTTGG ACCATTGGAG 1020

40 GAACCTAGTG TCACGCACAA ATGGGGCTAT TCCTACGCTT AGAATAGGGC TTGTCTGCCC 1080

ACTTTAGAAG AGTCCCAGGT TGGTGAGCAT TTACAGGGAA GCAGGGCAGA ACTCTGAACG 1140

45 ACAATACGTC TCTCTGAGCA GAGACCCCTT TCTCTCTGT ATCCACCCAT ATGGACTTGG 1200

AATCAATCTT GCCAAATATT TGGAGAGATT GTGTGGATT AAGAGACCTG GATTTTATA 1260

TTTTACCACT AAATAAAAGT TTTCATTGAT ATCTGTCTT GAAAAAAAAA AAAAAAAAAA 1320

50 AAATCGA 1328

55 (2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1856 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	GAATTCGGCA CGAGCTCTGC AACATCTGAA ATGAATTTG ASTGAGGGT TCCGCTGCCC	60
	CCTAGATTAA ATTCCCCGGG CTGAAATTGA GTTGCAGAT TAAATATCA TATTTTAAAT	120
10	TGCTGTCTTC AATTAAAGCA TTATAGACA TACTAATT TCAAGTTTC GATGCATGCT	180
	TTTCCAGGCC TTCTTTCTTT GTACAAAAT AATGTGCA AAGCTTTTC ACTTATATTC	240
	TTCAACATG ATGCTAATT AATTAATTA GTTCTAGA TATTTATTA TTCCTATGAT	300
15	TTTGGCACTG TTATTAAGTC TCTCAAAAT AATCTAGG AAGAGGATTA TTTTAAGTAA	360
	TTTGATTATC TTCTATCTC TTATATTAC TTCTACTTA CTACAGAAAT TCGTTCCATT	420
20	GCTTGGCATT GATACAGTAA ATTTGTAAT GAGGAGACA TAAAAAAAT CTAAATTACT	480
	TGTGCTTAAT GACTGTAGCA GAATSCCTTT TCTTAATC AGTGTGCTT TCTTGCAGTT	540
	TAGTTTGATA GATTTCGAG CTATGCTGTT TCGAGGAG TATTTGGCT GGTAGGAAGC	600
25	CAGGCTTCTT TGTCTCTGT TGTAGCTTGC ATGATGCTT CATAGGCTG ACAACGTAGC	660
	CGGAGATCAC AATTCAGGCC CTTCGCTTAC TTCTATTTT TTCTAGGTC AGAGAGCTTG	720
30	GCAGAACTG ACCTCATCTG GCAAGGCTTG CCACTGAGT GATCTTTTA TGCACCTCAT	780
	GTGTTCAGGA AGCCACAGGC CATATTTTAC TCTGAGAA AATTAAGAG GAAAAACCCC	840
	ACAAAGTATA ACAACCCCTT AAGATATAT TATTTTAA TGAATTAAT TTTTCAGTTT	900
35	ATACCATTGG CCAATTACAA GATAAAAATG TCAATTTCT TACAGATCC TTTGTTGACT	960
	TGTCTTTTCA TCTCTTGCTA TTATATTTG TCACTGTA TCAACAAAG CTTATTTGCT	1020
40	GAGGAAGGAC TTTGCTGCAC TTACTGTAGC ACATCAACA CTGGGAGGG TGGTGTTTAA	1080
	CTTTTAAAAA AATGTTATTC TGAATTAAT AATTAATG GCTTTTTTCA TGAATACAGC	1140
	GCCACCTTGC AAGGTTTAGT GAGATTATG GAATTTGAT AATTAAGGAG GAATTGCTGC	1200
45	TAGCTCCAAA AATTTGGGAA GCAAAATCTA GCGCAATG TTTTGAAGT TTGAATCTGA	1260
	TTAACAGATT TGCAATTGAA GTGACTGAG ACATTAAGT CAGATTTAG TTAATAATAG	1320
50	AAAGAGGAAT AAGACATCT TTCTCTCTA GAAAGATA CACTGAAAT AATAATCCTT	1380
	CCCACCTTCA TTGAGATCAG CTGTCTGTA AACCTGATA GATTGGAAT ATGATAAACA	1440
	TGATAATAGT GGTACTTTTG TAATTTGCT GGTGATTA ABAATATGT AAAGATGAG	1500
55	TTCACTTTT CTCTGACAT TCTATCTCT AGATGATG TACTCAAAAT TGGGAATTAT	1560
	AACGTGCTA ATTTTGTG TGTACCTG TGGCCTTT GTTTAATAC CCACAGTGTA	1620
60	ACATTAAT ATCACTAT GACATAAT TAACTAGA TATTTAATG ATAAATTTA	1680

GGGGTAAATG TTTACTTCAA AATGACTCCA TATTTCAAAT ATCTGTITAG ACTGTGAAGG 1740
 CCAAATAATT TTTAAGAAAA CATTGAAGA GTACTGTGTT TGCATTTGTG AATAATCTTA 1300
 5 CTCACAGCRA GTAAACGTAA TAAAGCCAA CATTTAAGCC AAAAAAAAAA AAAAAA 1356

10 (2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1558 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

20 TGGGTATCCA TTCTGNAAT TACTTTACTT AGGATAATGG CCTCCAGCTC COTCCAAGTT 60
 GCTGCAAAAG GTATTATTTT GTTCCTTTTT GTGGCTGAGT AGTATTCCTT GGTGTATATA 120
 TACCACATTT TCTTATCCA CTCATTGCTT GATGGGCAGT TAGGTTGGTT CCACATCTTT 180
 25 GCAATTGTGA GTTGTGCTGC TCCAGATATC ATCTTTAACT CCTTTGCCTT CTCCACATAC 240
 ATTTCCAAGT CCGTTTCATT CTACCTCCAA AATGTATCTT GTATCCATTC ATCTCTCTCC 300
 30 ATCTTCAATC TATTTCAATG CCCCATCATC TCTTGCATGG AGGAGTGTA TAATTGGCTA 360
 ACTGGCCTGT TCTTACATTT TAAATCAAA AGATGTGACA GGTGAAATGC CTATTTCACT 420
 GTCCATTGAT GGTCTGCTT ACACACCACC TGGCTGCCTG GTGTCCAGT GGCAGAGTTG 480
 35 AGCAGTGTA AAAAGACTGC TTGGCCCTTT ACAGGGAAAG CAGGTCCACT GTGGCCTGTG 540
 AGGACGAGAG CTCTGGGCAG GCTCGGACAC TGGCAGACCC TGGTCTGGC TGGCCAAGGC 600
 40 AGCAGGGTAT GTGTTTCGGG TCACTCACAG GGCTCAGCAC CACTCCTCAT GGCTTCCTTA 660
 CTGTTTCGGC AGAGGCTGAC CCGCGGCTGA TTGAGTCCCT CTCCAGATG CTGTCCATGG 720
 GCTTCTCTGA TGAAGCGGGC TGGCTCACCA GGCTCCTGCA GACCAAGAAC TATGACATCG 780
 45 GAGCGGCTCT GGACACCATC CAGTATTCAA AGCATCCCCC GCGTGTGTA CCACCTTTTGC 840
 CCACCTCTTC TGGTGCCTTC TCTTCTGTCT CATAGTTGTG TTAAGCTTGC GTAGAATTGC 900
 50 AGGTCTCTGT ACGGGCCAGT TTCTCTGCCT TCTTCCAGGA TCAGGGGTTA GGGTGCAAGA 960
 AGCCATTTAG GGCAGCAAAA CAACTGACAT GAAGGAGGG TCCCTGTGTG TGTGTGTGCT 1020
 55 GATGTTTCTT GGGTGCCTTG GCTCCTTCCA GCAGGGCTGG CCCTGCGAGA CCCAAGGCTC 1080
 ACTGCAGCGC GCTCCTGACC CCTCCCTCCA GGGGCTACGT TAGCAGCCCA GCACATAGCT 1140
 TGCCTAATGG CTTTCACTTT CTCTTTTGTG TTAATGACT CATAGTCCC TGACATTTAG 1200
 60 TTGATTATTT TCTGCTACAG ACCTGGTACA CTCTGATTTT AGATAAAGTA AGCCTAGGTG 1250

314

TTGTCAGCAG GCAGGCTGGG GAGGCCAGTG TTGTGGGCTT COTGCTGGGA CTGACAAGGC 1320
TCACGAAGGG CATCCGCAAT GTTGGTTTCA CTGAGAGCTG CTTCTGGTC TCTTCACCAC 1380
TGTAGTTCTC TCATTTCCAA ACCATCAGCT GCTTTTAAAA TAAGATCTCT TTGTAGCCAT 1440
CCTGTAAAT TTGTAAACAA TCTAATTAAA TGGCATCAGC ACTTTAACCA AAAAAAAAAA 1500
AAAAAAAAA AAANAAAAA AAAAGGGGGC CGCTCTAGAG GTCCAGTTA NGACCGCG 1558

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 948 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

TAAAAATCAT GCTCTGTACC ATCCTCACC G TAGTCATCAT CATCGCCGGC GAGACCACGA 60
GAACTACTGG GATCCCTAAA AACGCCCTG GTCCGGCCCC ACTGTGGCC CTTGGATCTC 120
CCAGGCTCTT TCTGCAGWCA TACCGCGGAC CCAATGGGCG CCTTGCACAC CGTTTCTGG 180
GGCCGTGAGA CTGGGATACA TCGTAAACTC CGCCTCCAGG GAACGTCTCG CTTKCGGAGC 240
AAGMTCCGAA TCCAGTTCTT CAGGAACCCC TCGAAAACCC ACACCCCCAG GGACGCCGCT 300
TTCCGGGATC CCGGSCAAAC GCGGACCCCT CAGTCGCTCC AGGCCCCCTC ACCCTCAAAG 360
TGTAGCGCCC CCAACCGAGC AACCTCGTTT TGGTCCCTAA AACCCCGCCT CCTCTATAAG 420
CACCGCCCCA GCTCTGACAA AACCCCGCCT CCAGGTCCGC AGGCTCCGCT TCTTTTCTTC 480
TCCGCGGGGT GATTCACTCC AGTGATGGG TTTGTGGCTC CAGGCTCCG CCACAGACGG 540
ACAGACCCCT CCTTTCTTC CGGCAAAAG ACCGAGCCCT GGGGTAGTAA GGGCCCCACA 600
CTCCTGTTTT TTGCAAGTAC ATTTTGTGCC YTCCTCCACC CAGGTATCTG CCTATTTTCT 660
TGCTAATCCC AGAACCTTTC CTTTGTCTTT TTTAAGGAC ATTTGGGAAG TTCCTGGTGT 720
AGGACCCCTC TCCCTGGGAT AAGAAACCTG CCTGTAAACG CTCTGTAAAT ACTCCCTTCC 780
ACCCATCCCA GCCCCTGGG AGCCGGGCG AAGGGAATCC AGGCTATGGA CCTCCCAAGT 840
CCCCGCTCCC CGCTCCCTC GCGGGCCCCG CTTGTCTCTG ATCTGTGTGT GAGTGTGTGT 900
GAACTTCTGA AAGACAATAT TAAAGAGACT TAGTTGAAAA AAAAAAA 948

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 990 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

10 GGGGAAGTGC AGTGACAGCA GGAGTAAGAG TGGGAGGCAG GACAGAGCTG GGACACAGGT 60
 ATGGAGAGGG GGTTCAGCGA GCCTAGAGAG GCCAGACTAT CAGGGTGCCG GCGGTGAGAA 120
 TCCAGGAGGA GGAGCCGAAA CAGAAGAGGG GCAGAAGACC GGCGCACTTG TGGGTTGCAG 180
 15 AGCCCTCAG CCATGTTGGG AGCCAAGCCA CACTGGCTAC CAGGTCCCTT ACACAGTCCC 240
 GGGCTGCCCT TGGTCTGGT GCTTCTGGCC CTGGGGGCGG GGTGGGCCCA GGAGGGGTCA 300
 20 GAGCCCTGCC TGCTGGAGGG GGAGTGCCTG GTGGTCTGTG AGCCTGGCGG AGCTGCTGCA 360
 GGGGGGCGCG GGGGAGCAGC CCTGGGAGAG GCACCCCTTG GCGGAGTGGC ATTTG/TGCG 420
 GTCCGAAGCC ACCACCATGA GCCAGCAGGG GAAACCGCCA ATGGTACCAG TGGGGCCATC 480
 25 TACTTCGACC AGGTCTTGGT GAACGAGGGC GGTGGCTTTG ACCGGGCGTC TGGCTCCTTC 540
 GTAGCCCTTG TCCGGGTGT CTACAGCTTC CGGTTCCATG TGGTGAAGGT GTACAACCGC 600
 30 CAAACTGTCC AGGTGAGCCT GATGCTGAAC ACGTGGCCTG TCATCTCAGC CTTTGCCAAT 660
 GATCCTGACG TGACCCGCGA GGCAGCCACC AGCTCTGTGC TACTGCCCTT GGACCTGGG 720
 GACCGAGTGT CTCTGCGCCT GCGTCGGGGG NAATCTACTG GGTGGTTGGA AATACTCAAG 780
 35 TTTCTCTGGC TTCTCATCT TCCCTCTCTG AAGGACCCAA GTCTTTCAAG CACAAGAATC 840
 CAGCCCTGA CAACTTTCTT CTGCCCTCTC TTGCCCCANA AACAGCANAA GCAGGANANA 900
 40 NACTCCCTCT GGCTCCTATC CCACCTCTTT GCATGGGAAC CTGTGCCAAA CACCCAAGTT 960
 TAAGAAAAAA ATAAACTGT GGCATCTCCA 990

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1603 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

GGTCGACCCA CGCGTCCGGC CCGCCGGCTC CGGAGCGGCT CTGCCTTCCC GAGCGGGGA 60
 CCGCGCCCTG GGGGAGGAGG GCGAACGACG CCGCGATGGC TCCCGGGCA CTCCCGGGT 120

	CCGCCGTCTT AGCCGCTGCT GTCTTCGTGG GAGCGCGCGT GAGTTCGCGG CTGGTGGCTC	180
	CGGACAATCG GAGCAGCGCG ACATTGCACT CCAGAACAGA GACGACCGCG TCGCCACCA	240
5	ACGATACTGG GAATGGACAC CCAGAATATA TTGCATACGC GCTTGTCCCT GTGTTCTTTA	300
	TCATGGGTCT CTTTGGCGTC CTCATTTCGC CAMCTNGCTT NAAGAAGAAA GGCTATCGTT	360
	GTACAACAGA AGCAGAGCAA GATATCGAAG AAGAAAAAGG TTGAAAAGWT AGRATGGAAT	420
10	GACAGTGTGA ATGAAAACAG TGACACTGTT GGGCAAATCG TCCACTACAT CATGAAAAAT	480
	GAAGCGAATG CTGATGTYTT AAAGGCGATG GTAGCAGATA ACAGCCTGTA TGATCCTGAA	540
15	AGCCCCGTGA CCCCCAGCAC ACCAGGGAGC CCGCCAGTGA GTCCCTGGGCT TTGTCACCAG	600
	GGGGGACGCC AGGGAAGCAC GTCTGTGGCC ATCATCTGCA TACGGTGGGC GGTGTWGTCC	660
	AGAGGGATGT GTGTCATCGG TGTAGGCACA AGCGGTGGCA CTTTATAAAG CCCACTAACA	720
20	AGTCCAGAGA GAGCAGACCA CGGCGCCAAG GCGAGGTGAC GGTCCCTTCT GTTGGCAGAT	780
	TTAGAGTNAC AAAAGTGGAG CACAAGTCAA ACCAGAAGGA ACGGAGAAGC CTGATGTCTG	840
25	TTAGTGGGGC TGAACCCGTC AATGGGAGG TGCCGGCAAC ACCTGTGAAG AGAGAACGCA	900
	GTGCCACAGA GTAGCAGGTG AGCCGTGGTT TTGGTGACAT TGGGGGCAGA GTGGTGCAGG	960
	GTGAGGAGAA GGTACTTGA GCCTCCCAGG TGCTGTGGCA GCATAGGAAT GGTATTTGAC	1020
30	AGGGAAGTGG GAGAGCTTTC CTTGACCCAG GAAGACTGAG GGGGACTGAA CATGATTACT	1080
	TGTCTGCCTA GAGCTTCTTG TAAAGAAGTC ACAAACCTAG TGCTCCAGG GGCTTGGCTG	1140
35	TGTGATAATG AGGATAGAGG ATTACTTGTG AGGCAATGTG GCATGGTGGG GATTGTGGCA	1200
	AACTAGAATT CACATCACCC ACCATATAGG GCTTGCATTA CCACGAGGCA GAAAGCACCT	1260
	AGTGTGCTG CATCTCTTA CGCAAAAAG ACAAAATCCA GACTTCTAAA ATGTAAAATC	1320
40	ACTGATTTTC GATATTGGCA GCTTACTTTT TTTTTTTAAA CAACCATGCA GGCCAAATGA	1380
	CTTGTAATCT TGTACCAATT TTTAGGTAAA CTGTGACTTG AAAAAGTCTG GAGCAAACAA	1440
45	ACCAATGCTT TTTCTTTTA TTCTGTGGR AACCACTTTT CTTTGTGTCA CAGTTYTGAA	1500
	ACCTCAATAC GAATATTTCT CTTCCACCA AATATTTTGA GGCAATTGAA AAGCCACAGT	1560
50	GATTTATTTT TTGATTGGC AATTTTAATT TTGCAAGACA ATT	1603

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1052 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

5 TACAGCTCAG GATGCCTGTA ACATTGTGAT CTCTGGGCTT CTGGGTCTTG CTTAGCCTGC 60
 TTTTTCCTTG GAGGACTGAC CAGGGATGCG GCGGAGCAAC ATGTTACTAA ATCATACTCT 120
 CCTCCCTACC TTTCCGAGAC CTCTCACTCC TGCCTGGTGT TCCAACCCGT TGTGTGGCCA 180
 10 GAGTATACAT TTTGGAACCT CTTCGAGGCC ATCCTGCAGT TCCAGATGAA CCATAGCCTG 240
 CTTCAGCAGN AAGGCCCCGAG ACATGTATGC AGAGGAGCGG AAGAGGCAGC AGCTGGAGAG 300
 GGACCAGGCT ACAGTGACAG AGCAGCTGCT GCGAGAGGGG CTCCAAGCCA GTGGGGACGC 360
 15 CCAGCTCCGA AGGACACGCT TGCACAAACT CTCGCCCCGA CGGGAAGAGC GAGTCCAAGG 420
 CTTCTGCAG GCGTTGGAAC TCAAGCGAGC TGACTGGCTG GCGGCTCTGG GCACTGCATC 480
 20 AGCCTGAATG AGGCTGGCCA CTTGCCACTT TGCCTGCCC TCTGCTCCA GGGCTCQCT 540
 MYCCTTCCTT TTCTTGGTGA AAGGCACCTC CTTCTCTGAT AATGAATGGT GTTCCCTTTG 600
 CTTGGCTGGG GAGCCCCCA GGCAGGTTT GCTGGCCATA GATACCTTTG GGCTGCCTGR 660
 25 GACAGGCTCC TGAGGAGGAT TGAGGGTGAA AGTCTCCAC GAGTACACTA AACCTAGGTC 720
 TGGTCACCAA TAGGGTTTGG AGAGCAAAGG GCCACAATC ATCAGCTGCC TGTCTCTTAG 780
 30 ATGCACTTTT TTTTCCACC AGCACATCCT TCAACACACA GAATTCAGG GAAGAGTTCT 840
 CCCCAAAACC CTAGCTCTTT ACCCTTCCAT TTAGCCTTC CACCCAGCTT CCACAAAAGA 900
 TTTGGCTCTA CTTGGATCT GCTAGTAAAT AACTAATAGG CAGGCAGTTA TTTGGGTAAG 960
 35 GAAAAAAGGG GTGGGAGAGA CAGAAATTT GCCCACTGCT GCTCCTCCCC TTGGSTVTC 1020
 ACCTGGGATT TGCTATTGAA TCTCTACCT NN 1052
 40

(2) INFORMATION FOR SEQ ID NO: 58:

- 45 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 814 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

ACNCGNTGGC GCGCGCTCTA GAACTAGGGG ANCCCCGGG CTGCAGGAAT TCGGCACGAG 60
 55 CATAGACTTT TAAACTGGTA CGGTCTTAG AGATGGTCCT TGGCCTCTG TTGTTGTGT 120
 KGTFTTFTT TTTTCTCT TCTCCTCTC CTTCTCTTC TCTCTCCTT CTTCTCTCT 180
 60 TTTTFTTCA GAGTCTTGCT CTGTCACCAA GACTGGAGTG AAGTGATGTG ATCTGGGCTT 240

ACTGCAACCT GGGAGGCAGA GGTTCAGTG AGTCGAGATG GTGCCATTCG TCTCGTTTGG 300
 GCAACAAGAG TGAACCTCTT GTCTCAAAAA AAAAAAAAAA ATGAGGTTTA AGACAGTTTT 360
 5 GTCATTACTG GTGGCATCTG GTCACACAAG ATAGCATTAA ACGTGACATG GCACATAAAA 420
 TTGGTTAAAA AATTTTGTTC TTAAATTACG TAATGTAAAA GCCCAACAAA CACTTTATGC 480
 AAGATTGGAA TGTATCTTCA AATTCAGATT TAATAACAT GTAAAGATCC TCTGTATATA 540
 10 AAAGTTGTAT TTAATCCCTT GTGCCCCAAG AATGCTATAA AAGATCCCAA GAATGTTATC 600
 TATGAAAAGA TAGCAATAGG GAATGCTGAA CAAATTAATT AATTTGCCAA TTCTAAAAAA 660
 15 CATGGACTTA AACCCCATGA AAACCTGGTT CCATAGTTTT AACTGTTTTA TGGTTCCAT 720
 ACAAACCAG AGTGGTTTAC ATTCCACAAT NACCAATTT GCATCCAATN TTGGGGTAAT 780
 20 TTNGGTATT TGCCATGGGA TACTATTCAT TTTC 814

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1215 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AGAGGAAGTC TTTTGCCAAG CCTGTTCTCT GGAATAACGC CATCCAGGCT GGGAGGGGAA 60
 35 GAGTGCTCTG CTACACTCGT CCCCCTCCTG CCTCATCTTC CTTCCTCAGC TTGGTTCTCTG 120
 ATGGGAACAG AATGGAGGGC CTGAGAACAT ACTTTCTAAA TGCCCTTTGAC CCAGGAACCG 180
 40 ATTATCTATA TTTGTCCCA TTTTCCTTCA CCGTGACATT CCAGCATTGT CTGACTGTGA 240
 GGTGGGCCTT TGAGAGCCTC CAGGTTCTCT AAAACAGGCC TGAGCGATGG GCATCACACC 300
 CTCTGCCTAC CCACRTGCCT GCTTACCTGC CAGATAACCA AGTGNAGATG TCTGCGAGTG 360
 45 GCTAGTTTTT ACATCTTTAC TAGTGTCTTG YTCACCTTTG GGCAAGGGCC CCTCTAGGC 420
 CTTGCCCCAC CTCCATCAAA CCGAGACACT GTAGTCAGAC CTCAGYAATA TAGGAGGCAA 480
 50 TAATCTTTTA ACACTGTTTT GCAACAAAC AAAAAGAGAA AAATCCCAGC CAGGGGAACT 540
 CGCCACCTGC CCACGCTAGT TCCATCCAG CTCAAGACCC GCCCTTAGAC CAGGCAGGCA 600
 AAGCCCCCA TCACACTCGG CCACTAGTGG GGTCTTGAGG CCAAGAAAGA AACCAGACCC 660
 55 TGTATGACAA GTTGGKTCCT TTCCAGAACA CGACAGAAAC AGGGGGGGCC CTTGTTAAT 720
 GCCACTCCAT ACTCCAGAAG CATTATTCCT TATTTGGGAC AGCCAAGGGC AGATTACAG 780
 60 GTTATTGTAG GAATAAGAC TAGTTTACAA AGGARAACA GSCCCTGGAC TTCCCMAGCA 840

AAGGTCAGGT TAGGGCTCCT GTACCCATTC TGTTCACCA CTGTTTGATC TCTCTGGCCT 900
 CCCACCAGCA ATGCCGTTTC CTTTTTATGG ATCTGTTGGG AACCGAGAG AATCAACAGA 960
 5 TCAATGACAT AGGATCCGAA GTGCAATGAT AGTCACCTCT AGTTTGGCAT TTCACAACT 1020
 CTGNACACCA AGGTATTGGT AGGTTACTCA ATTTCAAAAG GGCCCCATGG CCAAATATGT 1080
 10 TTAGGAACCG CTGTTTGNAT TTCTTTTTTT GGAGACCGAT TGTATATAAT ATATGTCAAA 1140
 GGCTTTCCGA ATTCCTGCAG GAAAGAAATC AGCTTTGTTA AATCCNAAAA AAAAAAAAAA 1200
 15 AAAAAAATAG ACTCG 1215

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 478 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ATTTCTTATG ACATGGGGGT TTGAATTGGT TGGCAATGT TTAATTTTAA TATCCATAAT 60
 30 CAGTGAGGTC CTGCTGGCTG TAATCATTAA TTGTGAAATC TAAGGAGCTT AGTTCATGGC 120
 TCTAGAATTT CACAGAAAAA TGYGMTATGA TACGAGCATT AAGTTTATTT CTCTGATCT 180
 35 TTGATGCAGC TTTGTTCACT TTATCTGTTT TTGTATTTAT TGGTCATCTA CTTCOCATGC 240
 CAAAAGGGAC TGGTCTACAT AGCTGCCGCTA AACACCTGAT CAAATCACTA AAAGAAAATG 300
 TGTACCTCT AATGAATTAT COTGATTGTA AGTTAAAAAT CAATATTTCC CCGTAGTGAG 360
 40 GTTTGCTTTT TAAAAAGAAK KCTTAAAAA AAAAAAATAA AAACGAGTTN AAGAAAAGGA 420
 AGCAAGCTCA GGTAAAGGTG ACACATTGGG CTAAGGAAGC TAGAGCCTGT GGAGANGC 478

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 618 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TATGACCTTG ATAACCCCAA GTTNGAAAT AACCTTCANT AAAGGGAACA AAGCTGGAG 60
 60 TTGCGCGCCT TGCAGTTGCA CACTAGTGGA TCCCAAGAA TTGGGCACGA GTCATATGA 120

320

GCTACTAGGT AAGCCTTCTG GGACTTTCAG ATATTTTGGG GAACATTGAT TTTTGTCTTT 180
 ACATGCTGTG GACCCTTGGC CATCAAATGG TATGGGAAG CTCATCCGTC TGTCTGTGAT 240
 5 GGTGATGTCA GTCAGGCGTC TTTTGTAGTAT TTAGTGGGTC CTCAGTACTG TGCCAGATGC 300
 TGTGGGGAGC CGTGGTGCTA TGGAGGAGGA GTGCTCCAGA GGACTCTGCT GTGTGGCAGG 360
 10 CCAGCATAAA CAAGCCAAAGG GAAAAGGCA GGCATGGAAT AAAGGGGGAG AATACCAAGT 420
 TGTGACTTAC TGCTGACTGT GTGGATTAGC CTATCAGCAG TAATCAAGCA GGGCGGAGGG 480
 CATTATCTTT GAGCCAGAAG AGTGAGCACT GGSCCGAGGG TGGAGCATCA AGAGGGGGTG 540
 15 TAGGACCINCA AGCCTTCTTN CNGGGGAGAC AACGTCAATA AGCNGTCAGT AGTCACCGAC 600
 AGTTTTGGGA AGCAAGGG 618

(2) INFORMATION FOR SEQ ID NO: 62:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 751 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TCGACCCAGC CGTCCGAGGA GCTGGACTTC TGAGACAGCC ATTCTGCTTG CATAGCACTG 60
 35 TCTGCTGCTA CAGCTCATAG AAGTCAACAA TTTTCTTCAA CACTGGTAGG CAGCCTCTAA 120
 ATGGCCCTGA TCACCTCAC CTCCTGCCAT TCACACCNNT GTAAAATTCC ACCCCTGGAC 180
 40 CTAGTGACTC ACTTCTAACA ANGAGAATAC AGCAAAAGTA ACATCGCTTC TGAGGTGAGG 240
 CTACAAGGAG ACTACGATGC CTGCCTGGT CAGCCTTCTC CTGCTCTTTC CATTGCTCCC 300
 TCTGATGGAA GCCAGTTGCC ATGTGATGAG GTGCCCTATG GAGAGGCCCA CGTGACAAGG 360
 45 TATTGTAAAA AGCCTCTGAC CAATAGCCAT CTAGAAACGG AGGCCAGTC CAGCAGCCTC 420
 TGAGATGAAT CCTGCCAACC TGAGCTTGGG GACAGATTCT CTCCTATCC TGCCTTGGGA 480
 TGATCAGAGC CACCACCAAC ACTTTCACCT CCTGGTGAGA GGCCAAGCCA GTGAACCCAA 540
 50 GGTAAACTGG ACAGAATCCT GACCCACAGA AACTGAGATA ATGTTTGTTA TTTTAAGCTG 600
 CTCAGTTTGT TACAGAGCAA TAGATRACTA ACTCAAACAC CATAAAATTC TAATATTTTA 660
 55 TTCTATCACA CAAACCAGGT AATACCAAGT AAATGCCATT ACTATACACA TATTTTGTGA 720
 ACACAATTAC ATGTGATTTT TTAAGAAGGC T 751

60

(2) INFORMATION FOR SEQ ID NO: 63:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 780 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CAGGCACTCA CAGTCCCGGA TCCCGGGTTC GACCCACGCG TCCGGGTTGG CAACTCCTGA 60
GGCTTGACAG GGTGACTTCA CATTTTCCTA CTTCTGCTTC TAATCTCTTC TAGAGCACCT 120
GCTATCCCGA ACTTCTAGAC CTGCTCCAAA CTAGTGACTA GGATAGAATT TGATCCCGTA 130
ACTCATTTTC TCCGGTCTTC ATGCTGCTA ACAGCAATTC CTGTGCTCTC CTCTCAGGGG 240
CAGCAGCTTA ACCGGGGGAC GTGCTAATCC AACTGGGAGA AGCCTCAGTG GTGGAATTCC 300
AGGCACTTTC ACTGTCAAGC TGGCAAGGCG CAGGATTGGG GGAATGGAGT TGGGGCTTAG 360
CTGGGAGGTC GTCTGAGCA GACAGCGAAT GGGAGAGGAG GATGGGAAGT ACACACTGGC 420
TGTATGGCT CTGAGGCTCC CTGGGGGCTG CTGAGCTCC TCTGCTCTCT TGCTGTTTTT 480
TGATGATTTG GGGGCTTGGG ATCCCTTTG TCTCATCTG AGACTGAAAT GTGGGGATCC 540
AGATGGGCT TCTTCTCTCT TACCCCTCTCT CCTCAGGCT GCAACCTCTA TCTTGGAAAC 600
TGTCTTCTCT TCTTCCCAA CTATGCACT GTGTCTGCT CCTCTGCAA GGCCAGCCAG 660
CTTGGAGCA GAGAGGAAT AACAGCAAT TCTGATGCCA AAAAAAAAAA AAAAAAAACC 720
GGGGGCGAAA GTTATATNCC CTTAAGTAA GGGGTAATT TTTAGCTTGG GCACTNGGCC 780

(2) INFORMATION FOR SEQ ID NO: 64:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 588 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

TTCCGATTA ATCGACTAC TATAGGAAT GCGCTCGCCA TGACCCGCGG TAACCAGCGT 60
GAGCTCGCCC GCGGAGAGA TATGAAAAG CAGAGCGACT CGGTAAAGG AAAGCGCGA 120
GATGAGGGG TTTCTCTTC CCGCCGCAAG CAGAGGGACT CGGAGATCAT GCAGCAGAAG 130
CAGAAAGG CAAACGAGA GAGGAGGAA CCGAAGTAGC TTTGTGGCTT CGTGTCCAAC 240
CCTCTTGGCC TTGGCTGTG TGGCTGGAGC CAGTCCGACC ACCGTGCGCT TTCTCTCTGT 300

322

AGTGCTCACA GGTCCCAGCA CCGATGGCAT TCCCTTTGCC CTGAGTCTGC AGCGGGTCCC 360
TTTGTGCTT CTTCCCTC AGGTAGCCTC TCTCCCTG GGCCTCCC GGGGTGAGG 420
5 GGGTTACCCC TTCCAGTGT TTTTATTCC TGTGGGGCTC ACCCCAAAGT ATTAAAAGTA 480
GCTTTGTAAT TCCAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 540
AAAAAAAAA AAAAAAAAAA AAAANNCGGG GGGGGGCCCC CCCCCCCC 588

(2) INFORMATION FOR SEQ ID NO: 65:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 774 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

TTTAAAGATG AAGAAATGAC AAGGGAGGGA GATGAGATGG AAAGGTGTTT GGAAGAGATA 60
25 AGGGGTCTTA GAAAGAAATT TAGGGCTCTG CATTCTAACC ATAGGCATTC TCGGGACCGT 120
CCTTATCCCA TTTAATTAAT TTCTCTGACA ATTCAATTAT TTTCTGTTAT TAATGTTGCC 180
30 ACTGCTTTCT GTTTGTCTGC ACTTCTTGA TAAATATTG CTATCGTTT ACTCCAGTCA 240
TCCGATGTTG CTGAGATTTA CATATGACTC TTGTCAACAT CTCATCTTT GACCGAATCT 300
TATTCATTTA ATAAGAGGTC TCATTCATTT GCATGGAAAA ATGCTCATTC TATATTGCAA 360
35 AGTGAAAATA ACGAGTTGCA AAACAGTGTA TACATATATG TGTGTATATA TGTACACTTT 420
ATTGTACAT TTCTATGTGA CATAATGCAA AGGAAAGTGT CTGATTTTAT TATACACCAA 480
40 AGGTTAACAG TGAATCTCTG TGTGATCTCT TTTTCTTCT TTTGCCTAT CTGCATCTTC 540
TCACTTGCCA AAAAATGAAT ATATGTTTAT GTGTGTATAT TACTTGTGTC ACAAAAAACC 600
CTAAAGTAGA CAGTAAAGA ACTTGTCAAT CGCCTTTGGA AGGCAATGAA ACACTTAATA 660
45 AACTCTCAAT AACAGAAGCG TAAAAATGAA ATGTAAACCT CCAATTACCT CTGGATCTCT 720
TAGCCAGAGT AATAAACTGG TAATTATTAC AGATAAAAAA AAAAAAAAAA AANA 774

(2) INFORMATION FOR SEQ ID NO: 66:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1866 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	ACCCACGGGT CCGGTCTCT TCTTCAGCAC ATGCCAAAGC TGTTCTCTAC GGCTGTGAG	60
5	ACAAGAGCAT CTTGGATGTA GCACAATGGA AGACTTAGAT GCCTTATTGG AGGAACTGGA	120
	ACGCTCCACC CTTCAGGACA GTGATGAATA TTCCAACCCA GCTCCTCTTC CCTTGGATCA	180
10	GCATTCCAGA AAGGAGACTA ACCTTGATGA GACTTCGGAG ATCCTTTCTA TTCAGGATAA	240
	CACAAGTCCC TTGCCGGCGC ATTCGTGTAT ACTACCAATA TCCAGGAGCT CAATGTCTAC	300
	AGTGAAGCCC AAGAGCCAAA GGAATCACCA CCACCTTCTA AAACGTCAGC AGCTGCTCAG	360
15	TTGGATGAGC TCATGGCTCA CCTGACTGAG ATGCAGGCCA AGGTTGCAGT GAGAGCAGAT	420
	GCTGGCAAGA AGCACTTACC AGACAAGCAG GATCACAAGG CCTCCCTGGA CTCAATGCTT	480
20	GGGGGTCTSG AGCAGGAATT GCAGGACCTT GGCATTGCCA CAGTGGCCCA GGGCCATTGT	540
	GCATCCTGCC AGAAACCGAT TGCTGGGAAG GTGATCCATG CTCTAGGGCA ATCATGGCAT	600
	CCTGAGCATT TTGTCTGTAC TCATTGCAAA GAAGAGATTG GCTCCAGTCC CTCTTTGAG	660
25	CGGACTGGCT TCGNCTACTG CCCCACGAC TACCACCAAC TTTTCTCTCC ACGCTGTGCT	720
	TACTGCGCTG CTCCCACCTT GGATAAAGTG CTGACAGCAA TGAACCAGAC CTGGCAGCCA	780
30	GAGCACTTCT TCTGCTCTCA CTGCGGAGAG GTGTTTGGTG CAGAAGGCTT TCATGAGAAG	840
	GACAAGAAGC CATATTGCCC AAAGGATTTC TTAGCCATGT TCTCACCCAA GTGTGGTGGC	900
	TGCAATCGCC CAGTGTGGA AAACCTACCTT TCAGCCATGG ACACTGTCTG GCACCCAGAG	960
35	TGCTTTGTTT GTGGGGACTG CTTCAACAGT TTTTCTACTG GCTCCTCTCT TGAAGTGGAT	1020
	GGACGTCCAT TCTGTGAGCT CCATTACCAT CACCGCCGGG GAACGCTCTG CCATGGGTGT	1080
40	GGGCAGCCCA TCACTGGCCG TTGTATCAGT GCCATGGGGT ACAAGTTCCA TCCTGAGCAC	1140
	TTTGTGTGTG CTTTCTGCCT GACACAGTTG TCGAAGGGCA TTTTCAGGGA GCAGAATGAC	1200
	AAGACCTATT GTCAACCTTG CTTCAATAAG CTCTCCCCAC TGTAAATGCCA ACTGATCCAT	1260
45	AGCCTCTTCA GATTCTTAT AAAATTTAAA CCAAGAGAGG AGAGGAAAGG GTAAATTTTC	1320
	TGTTACTGAC CTTCTGCTTA ATAGTCTTAT AGAAAAAGGA AAGGTGATCA GCAAATAAAG	1380
50	GAACTTCTAG ACTTTACATG ACTAGGCTGA TAATCTTATT TTTTAGGCTT CTATACAGTT	1440
	AATTCTATAA ATTCTCTTTC TCCCTCTCTT CTCCAATCAA GCACTTGGAG TTAGATCTAG	1500
	GTCTTCTAT CTCGTCCCTC TACAGATGTA TTTTCCACTT GCATAATTCA TGCCAACACT	1560
55	GGTTTTCTTA GGTTCCTCCA TTTTCACCTC TAGTGATGGC CTAATCATA TCTTCTCTAA	1620
	TTTGGTCTG ATACTTGTTT CTTTTCACGT TTTCCCATTT CCTGTGGCT CACTGTCTTA	1680
60	CAATCACTGC TGTGGAATCA TGATACCACT TTTAGCTCTT TGCATCTTCC TTCAGTGTAT	1740

TTTTGTTTTT CAAGAGGAAG TAGATTTTAA CTGGACAACT TTGAGTACTG ACATCATTGA 1300
 TAAATAAACT GCCTTGTGGT TTCATATAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1360
 5 AAAAAA 1366

10 (2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1152 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

20 CTCAAGGATG TAAAGGCTCT GCAGATTTTC GGAGGCCTGT CTCCCAGCAC CTGATGGGAC 60
 ACTTTTGGCC CCACTGTAAA TTCTGGGTGT ATCCTCCACT GTATGCTGTC ACCCCAAGGG 120
 25 CAAGCACTGC ATCTGCTTAG TGAAGGATTT ATTGTTCCGA AGATACATTT TCCCCTTKAG 180
 CAGAGACTGG CGTATCCTGG CAGTCTTCGG TGAGCCAGTT GTACCAGGAT TATGAAATGC 240
 AGATGTTTAC TGTGTCTATT TTGCTGTCTAT TGCTACTGAG GAGTACTGAC CAGAATCATC 300
 30 TGCAACTVTT AGTTGGCAGA GAGGACCACT ATGGCCGGTA GCTCTTTTCT TTCCTGCCAT 360
 TGTGGGGATG ATTCCAGGCC AAAGATGATG GAAAGTATG GAAATCATCT GAAAGGTTGA 420
 AGCTTGGCAC GTGAAGCCAT TCATGACTTT GTAAGGCAGT TTTGCTGAAG GCCAGTTCTG 480
 35 CCCTGGGAGG GACGGAGGTG AATCCTCCTG AGTACCTGTG GTTTTCTTAC TTCCTGCTGA 540
 ATTTACCTAA GTGCCTGTTG TTTGCTTGCT GTGGAGCCTT TCTGCTATTT CATTTGAGGT 600
 40 GCAGATGCCT TCACTTTCCC ACCRAAAAAA CCCCMACCAA ACCTAAGACC TTAATGCAAC 660
 TAAGTYTNCC AAGTACTTTT TAACCCAATG GGATGAACAG CCTGTGGTCT GCTCAGATCA 720
 CCCTGAGTGC GTGTGAGAAG GCMTNGGCTT TGCCAGGAAA TCCAGGAAGG CAGGGCCGGG 780
 45 CTGTGTTGGA AGCTGGCTTA GCTGGTGGGG CAGCCTTATT TCAATTAAAA GGGCATTGAC 840
 TGGGAGCAGC AGTCCTGGAG TTTGTTGCAT TTCCTATTGC CCTCAAAATG AGAAACCAGG 900
 50 AAAATAGCAG ATTGGAGCCT TCGAGAAGGC AGTAAATGGC TGTTTTTTATT GACAAAAGGA 960
 AAACATTTTA CTGCCATCTC ACTGATGGCA TCTCACTGAC TTAATATGAA GGCANGTTGT 1020
 55 AGTAAAAAAA AAGTCTACA TTTTTCACC GCCACGTTCT TATATCCTGT TTGTCAGCCA 1080
 CTGCTCANAA GGSCATGTTG TCTTGGGGAN TANAGGCGCT CTCCTTCCCT CGTTTTCCTT 1140
 ATAGGTTGGG TG 1152

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2483 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

5	ACCAGGCGGT GCGCTGGGGG CGGGAGCAGC GCGKAGCCCG GCTCGGCCAC ACCGATCGCC	60
15	CGCCGCCATG GGCTCCTCGC AAAGCGTGA GATCCCGGGC GGGGGCACCG AGGGCTACCA	120
	CCTTCTGCGG GTACAAGAAA ATTCCCCAGG ACACAGAGCT GGTTTGGAGC CTTTCTTTGA	180
20	TTTTATTGTT TCTATTAAATG GTTCAAGATT AAATAAAGAC AATGACACTC TTAAGGATCT	240
	GCTGAAASCA AACGTTGAAA AGCCTGTAAA GATGCTTATC TATAGCAGCA AAACATTGGA	300
	ACTGCGAGAG ACCTCACTCA CACCAAGTAA CCTGTGGGGC GGCCAGGGCT TATTGGGAGT	360
25	GAGCATTCGT TTCTGCAGCT TTGATGGGGC AAATGAAAAT GTTTGGCATG TGCTGGAGGT	420
	GGAATCAAAT TCTCCTGCAG CACTGGCAGG TCTTAGACCA CACAGTGATT ATATAATTGG	480
30	ACCAGATACA GTCATGAATC AGTCTGAAGA TCTATTCAGC CTTATCGAAA CACATGAAGC	540
	AAAACCATTG AAACGTGATG TGTACAACAC AGACACTGAT AACTGTGAG AAGTGATTAT	600
	TACACCAAAT TCTGCATGGG GTGGAGAAGG CAGCCTAGGA TGTGGCATTG GATATGGTTA	660
35	TTTGATCGA ATACCTACAC GCGCATTTGA GGAAGGAAAG AAAATTCTC TTCCAGGACA	720
	AATGGCTGGT ACACCTATTA CACGTCTTAA AGATGGGTTT ACAGAGGTCC AGCTGTCCTC	780
40	AGTTAATCCC CCGTCTTTGT CACCACCAGG AACTACAGGA ATTGAACAGA GTCTGACTGG	840
	ACTTTCTATT AGCTCAACTC CACCAGCTGT CAGTAGTGTT CTCAGTACAG GTGTACCAAC	900
	AGTACCGTTA TTGCCACCAC AAGTAAACCA GTCCCTCACT TCTGTGCCAC CAATGAATCC	960
45	AGCTACTACA TTACCAGGTC TGATGCCCTT ACCAGCAGGA CTGCCCAACC TCCCCAACCT	1020
	CAACCTCAAC CTCCAGCAC CACACATCAT GCCAGGGGTT GGCTTACCAG AACTTGTAAG	1080
50	CCCAGGTCTG CCACCTCTTC CTTCATGCC TCCCCGAAAC TTACCTGGCA TTGCACCTCT	1140
	CCCCCTGCCA TCCGAGTTCC TCCCGTCATT CCCCTTGGTT CCAGAGAGCT CTTCTGCAGC	1200
	AAGCTCAGGA GAGCTGCTGT CTTCCTTCCC GCCCACCAGC AACGCACCT CTGACCTGC	1260
55	CACAACACT GCAAAGGCAG ACCGTGCCTC CTCACCTACT GTGGATGTGA CGCCCCCAC	1320
	TGCCAAGGCC CCCACCACCG TTGAGGACAG AGTCGGCGAC TCCACCCAG TCAGCGAGAA	1380
60	GCCTGTTTCT GCGGCTGTGG ATGCCAATGC TTCTGAGTCA CCTTAACCTT GAACCATTC	1440

326

TTGGAATTGG CGTGGTATAT TTAACCACGG GAGCGTGTCT GGAAACGCCA ACTATCATTA 1500
 ATTTCACTACT AGTTTGTACC GTATCTGTAG GCATCCTGTA AATAATTCCA AGGGGAAAAC 1560
 5 TAAACGAGGA CGTGGGTTGT ATCCTGCCAG GTTGAGTGGG GCTCACACCG TAGGGTGAGA 1620
 TGTGAGAAAG CCCTTGTATT TTAACAACCC AAAAAGAATT GTAAGGGTGG CTTCCTGCCA 1680
 GGCTTGCCTT GCGGTTCTTG GGGGTGTGCA TCTTCGGGAA AGGTGGTGGC GGGGCGTCCA 1740
 10 CTAGGTTTCC TGTCCCCTGC TGCTCCTTCC GTAAGAAAAT GAAATATTCT ATGCCTAATA 1800
 CTCACACGCA ACATTTCTTG TACTTTGTAA GTCGTTTGG AGAATGCAGA CCACCTCACT 1860
 15 AAACGTAAA CCGTAAAGAG ATTTTACTT TTGCTCTCCG TGAGTCCCAT CTCTACTAAG 1920
 GTTTACACAG GAATCCACC TGAAGACTTG TGTAAAGTT CTACAGCGCG CACTGTTAAC 1980
 TGAACGTCTT TTTCTTCAGC CTATACGCGG ATCCTTGTTT TGAGCTCTCA GAATCACTCA 2040
 20 GACAACATTT TGTAACTGCT GCTGTTGCTT TCTACATACA CTTATAAAG TGACATTTCA 2100
 AAAGAAATRA GGTGCCACAG TTTTAAACCA GAAGGTGGCA CTCTGTGGCT CCTGTAGTA 2160
 25 TTATAGCTAT ACTGGGAAAG CATAGATACA GCAATAAAGT ACAGTAATTT TACTTTTTTT 2220
 CTTGTGTTAC ATCTAAATTA CAACCCCTTA TTGCCACGTG TGCACCTACT ACTCTCCAGT 2280
 ATGTCTTATT ACTCTCCAGT ATGTCACGCA TCTTTAACTT TTCACGTGCT ATGTTTGCTT 2340
 30 TCTCCCATTT TTAAGAGATG GTAAGTTAAC TGGAATTGAT TTAAGTAATG AAATTAATG 2400
 CAGATATCCC TGTTTTGTAA ATAAAAA AAAA AAAA AAAA AAAA 2460
 35 AAAAAA AAAA AAAA AAAA 2483

40 (2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 536 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

50 GAGAAATGGA GCTTGTGTAG ATAAAAATTT TTTCACGCA AACAGTCATT TTCCAGTGAA 60
 AGGAGAGCGT ATCCGCCGTA GGATGGACTT AGATCGTSTA AAAGCTGAGG CCACCCAGGA 120
 55 TATAACCTCC GGGGTCTTTT GCCTCCTTTT CTTAGACTC CCTCCAACT CGTGTATCTT 180
 TCCTTCAGCA GTACTGGGCT CCACGCCAAC CTAGTCCTTT GTCTTTACCC TATTACCTTT 240
 CATAACATCC TAGTTGAAAA GTATTATTC AACCGGTTT GAAATGAGA ACAGGTTCC 300
 60 AGARGCTAGG TTAATTGCGA AGGTGTTCA ATTAGTAACG AGTAACGCCA GCACTGCCAG 360

327

TTTCTTGCTT CGGAATTCTC ATGGTAGCTT TCACCARGCT CCCCCTCAA TGCTAACGTC 420
AACTACTGAA CTAGATTAGC AAAAAGGTCT TTAAACAGAA TTCTTGTTTT TCAGACAGAG 480
5 TTTCTTTTCAT GAAGCGCCCC ATTTCTACAG AGGAAAATAA ACTCCAAGCA GCCAGT 536

10

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCACGCGTCC GGCCTTTCTT GGCCAGAGGC GCGGTTGGA CTCACGGCGG GGGCATGATG 60
GGTAACAGGA CCGGTGGGGT CCCCAGGAAG TCCTAGAGGG GGTCCGGGTT TGGGTGGACA 120
25 AGCTTTCTCT GTCTCTCCC GACAGAGCTG ACCTGTCTG GGTTCACCG GGAGCGGGCA 180
TTTCCACCGG ACCGGAGCGT TCGGGGTGTC CCGGGCTGGG GAATACGTAG GGGTTGCCCG 240
GCGGTGTGGG GAGTTGGGGC GTGTGGCTCC AGTCCCGGGA GTTCTTGGAG GGGGTCCGCC 300
30 CACCGAGCTT CCGGACCGGC TGATCTGCCC GTAGCTTGCC GGANGGARGG CGGAGCTGAC 360
TCTCCGTCCC TTCTCCCATC CCTCCAGTG GTGGGTACGG GCACCTCGCT GCGGCTCTCC 420
35 TCCTCTCTGT CCTGTGCTT CTCTGCTGGG ATGCAGATGT ACAGCCGTCA GCTGGCCTCC 480
ACCGAGTGGC TCACCATCCA GGGGGGCCCTG CTTGGTTCCG GTCTCTTGGT GTTCTCGCTC 540
40 ACTGCCTTCA ATAATCTGGA GAATCTTGTC TTTGGCAAAG GATCCAAGC AAAGATCTTC 600
CCTGAGATTC TCCTGTGCTT CTTGTTGGCT CTCTTTGCAT CTGGCCTCAT CCACCGAGTC 660
TGTGTCACCA CTTGCTTCAT GTTCTCCATG GTTGGTCTGT ACTACATCAA CAAGATCTCC 720
45 TCCACCTGT ACCAGGCAGC AGCTCCAGTC CTCACACCAG CCAAGGTCAC AGGCAAGAGC 780
AAGAAGAGAA ACTGACCCCTG AATGTTCAAT AAAGTTGATT CTTTGTAAAA AAAAAAAAAA 840
50 AAAAAAAAAA AAAAAAAAAA AAAAA 865

55

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 932 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

5 TCATCATATA CAAAGTTTTT CGTCACACTG CAGGGTTGAA ACCAGAAGTT AGTTGCTTTG 60
 AGAACATAAG GTCTTGTGCA AGAGGAGCCC TCGCTCTTCT GTTCTTTCTC GGCACCACCT 120
 GGATCTTTGG GGTTCCTCCAT GTTGTGCAAG CATCAGTGGT TACAGCTTAC CTCTTCACAG 130
 10 TCAGCAATGC TTTCCAGGGG ATGTTCAATTT TTTTATTCCT GTGTGTTTTA TCTAGAAAGA 240
 TTCAAGAAGA ATATTACAGA TTGTTCAAAA ATGTCCTCTG TTGTTTTGGA TGTTTAAGGT 300
 AAACATAGAG AATGCTGGAT AATTACAACCT GCACAAAAAT AAAAAATCCA AGCTGTGGAT 360
 15 GACCAATGTA TAAAAATGAC TCATCAAAAT ATCCAATTAT TAACTACTAG ACAAAAAGTA 420
 TTTTAAATCA GTTTTTCTGT TTATGCTATA GGAAGTGTAG ATAATRAAGT AAAATTATGT 480
 20 ATCATATAGA TATACTATGT TTTTCTATGT GAAATAGTTC TGTCAAAAAT AGTATTGCAG 540
 ATATTTGGAA AGTAATTGGT TTCTCAGGAG TGATATCACT GCACCCAAGG AAAGATTTTC 600
 TTTCTAACAC GAGAAGTATA TGAATGTCTT GAAGGAAACC ACTGGCTTGA TATTTCTGTG 660
 25 ACTCGTGTG CTTTGAAC TAGTCCCTA CCACCTCGGT AATGAGCTCC ATTACAGAAA 720
 GTGGAACATA AGAGAATGAA GGGGCAGAAT ATCAACAGT GAAAAGGGAA TGATAAGATG 780
 30 TATTTTGAAT GAAGTGTGTT TTCTGTAGAC TAGCTGAGAA ATTGTGACA TAAAAATAAG 840
 AATTGAAGAA ACACATTTTA CCATTTAAAA AAAAAAAAAA ACTNGAGGGG GGCCCCGTAC 900
 35 CCAAATCGCC GCATAGTGAT CGTAAACRAAT CT 932

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 996 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

50 CGCCTGGCAC CATGAGGACG COTGGGCTC TGCCTGTGCT GCTGCTGCTC CTGGCGGGAG 60
 CCCCCGCGCG GCGGCCCCACT CCCCCGACCT GCTACTCCCG CATGCGGGCC CTGAGCCAGG 120
 AGATCACCCG CGACTTCAAC CTCCTGCAGG TCTCGGAGCC CTCGGAGCCA TGTGTGAGAT 180
 55 ACCTGCCCCAG GCTGTACCTG GACATACACA ATTACTGTGT GCTGGACAAG CTGCGGGACT 240
 TTGTGGCCTC GCCCCCGTGT TGGAAAGTGG CCCAGGTAGA TTCCTTGAAG GACAAAGCAC 300
 60 GGAAGCTGTA CACCATCATG AACTCGTTCT GCAGGAGAGA TTTGGTATTC CTGTTGCATG 360

ACTGCAATGC CTTGGAATAC CCAATCCCAG TGACTACGGT CCTGCCAGAT CGTCAGCGCT 420
 AAGGGAAGTG AGACCAGAGA AAGAACCCTAA GAGAACTAAA GTTATGTCAG CTACCCAGAC 480
 5 TTAATGGGCC AGAGCCATGA CCGTCACAGG TCTGTGTTA GTTGATCTG AACTGTTAT 540
 GTATCTCTCT ACCTTCTGGA AAACAGGGCT GGTATTCCTA CCCNGGAACC TCCTTTGAGC 600
 ATAGAGTTAG CAACCATGCT TCTCATTCCT TTGACTCATG TCTTGCCAGG ATGCTTAGAT 660
 10 ACACAGCATG TTGATTGGT CACCTAAAAA GAAGAAAAGG ACTAACAAGC TTCACTTTTA 720
 TGAACACTA TTTGAGAAC ATGCACAATA GTATGTTTTT ATTACTGTT TAATGGAGTA 780
 15 ATGCTACTTT TATCTTTCT TGATAGAAAC CTGCTTACAT TTAACCAAGC TTCTATTATG 840
 CCTTTTTCTA ACACAGACTT TCTTCACTGT CTTTCATTTA AAAAGAAAT AATGCTCTTA 900
 AGATATATAT TTTAYGTAGT GCTGACAGGA CCCACTCTTT CATTGAAAGG TGATGAAAT 960
 20 CAAATAAAGA ATCTCTTCAC ATGAAAAAA AAAAAA 996

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 785 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

35 GGCACGAGGG GCTTTGCGTA CACAATAGCT GCTAGGAGTA CCCAAAGCCT GARTACARCC 60
 TGCTGGTGTG ATGGCCAGT GTGAGCAGG CAGCGTCAMA CGGCTCGCTG TGACCCGTCC 120
 40 CGRAGACTGA AATGGGCCTG GGTCTTCTCC TKGCTCTGTG ATWAAAGTCC TCTCTTGAA 180
 GTGGAGAGCA AAGGCACACA GAGGTGCGG CTCACAAGAA TTCCTCCCGG TGACTGGGTA 240
 ATCAATGTTA CTGCTGTTTC CTTTGACAGG AAGACCACAG CAAGATTCTT TCATTCTCT 300
 45 CCTCCTAGCC TGGGGACCA GGTGGAAGT GACCTGGAC ATCAAAGGAG GGATTATGTG 360
 GCTGCTAAAG CCATCGGCCC ACAGCCTGT TCACTCTTG GTGCTTCTCT TTCCAGAGG 420
 50 CTGGTCCCAG CCAGGCACAC ACAAAGGCA GATTCTCGTA AACSCAGCCT CCTCCCTGG 480
 AGGCTGCCTC CTGCCCTGGA TGTGGAGTGG AGCTGCTCTG AGATTTTGAG TTCTCTGCA 540
 GAGATGATTA AATATATCCA AGAGACATTG GAAACCTGC TGAACATTTT ACATTGGTCT 600
 55 GCTCAGCACA TGGCTGGATG CGGATATTTC TATAATTCCA GAAAGTCACA CAGCTCCTCT 660
 GTATGAGACC AGTGGGCGCC ATTAAAAA ACAGGATGAG AATCTAAGAT ATATTATTAA 720
 60 TAAATGTAAT GGATTTTTTT TTTGTAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 780

AAAAA

785

5

(2) INFORMATION FOR SEQ ID NO: 74:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1069 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

20

TCCTCACCAT TCCCTAGGN CAGGTCCTTG CAGGTCCCAC ACTTCTCCCA GGTCCCTAAA	60
CTTGGGTGGG TCCTTTCCCT GGAGTAGCTG GNTCCTCCAG TCGAGGTCCC TGTTCAGTGG	120
GTTCTTAGGC TCCTGCACAT GAAGGTGTGT GCCTGTGGTG TGTGGGCTGC TCTAGGAGCA	180
GATACAGGCT GGTATAGAGG ATGCAGAAAG GTAGGCCAGT ATGTTTAAAGT CCAGACTTGG	240
CACATGGCTA GGGATACTGC TCACTAGCTG TGGAGGTCCT CAGGAGTGGG GAGAATGAGT	300
AGGAGGGCAG AAGCTTCCAT TTTTGTCTT CTAAGACCC TGTATTGTGT GTTATTTCCT	360
GCCTTTCCGA GTCCTGCAGT GGGCTGCCCT GTACCTGAA CCTCATGAGC CTCTAAGGGA	420
AAGGAGGAAC AATTAGGACG TGGCAATGAG ACCTGGCAGG GCAGATACA AGCCCAGCAC	480
CAGTGTCCCA GCCTTACTGG GTCCTTACCC TGGGCCAAAC AGGAGGGCT GATACCTCT	540
TGCTCTTCCT AGATGCCAC CTCCTACAAT CTCAGCCAC AAGTCCTCTC CACCCTAGGG	600
GGCTTGCTGC ATGGCAATAA CTCATAATCT GATTTGGAGG TTTGCCCTTT ACAGGGGCAG	660
ATTTTCTGCT CAGTTCAACA ATGAAATGAA GAGGAACTCC CTCTTTCTAC AGCTCACTTC	720
TATCAGAGGC CCAGGTGCT CAGAGCCACA TTGAGTTGCT TTTTCTGGGA TGAGGAAGTA	780
GGGTAAACT CCCAGTTTC CTGAGGGAGG CTCCTGACAG GTGCCCTTTG TCAGACCCTA	840
CCACAGCCTG GATAGGCAGC CACATTGGTC CTCGCCCTTG CTCGGNACTC CGTGGTGGTC	900
CTGCCCTTCT CCTGCAATC CTGTGGGTCT GCTCTGGTGT GTGAAGGTGG GTGGGTAAAC	960
TGTGTGCTA CTGAACCTGG CAAATAAACA TCACCCTGCA AAGCCAAAAA AAAAAAAAAA	1020
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1069

55

(2) INFORMATION FOR SEQ ID NO: 75:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 831 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

5 GGACATTAGA TCACTGTGGA CCTAAAACAA ACRAACAACAT ATAAGGAAAA TGSCATTAGA 60
 AATGGTCTGG GGATCAGTTT ATCACTGCAG TTGTACATC ACCCCATGGT CTAAATATCA 120
 10 GAGCTTTAGT CTGTCTCTGT TTCAGTTTAT TTTACAGGAG CTGAACATCA CACTTCCAGA 180
 AAAGTCTGTC TGGTATGAAA GGTATAAATT TGATATTCTT GTCTTTCACT TGAATGGCCA 240
 GTTCTGTATG ATGCATCCAG TAAACACCTC AAAACTTGAA AAACAGCTCC TGAAACTTGA 300
 15 GCAGCAAAGT ACTGGARGCT GACTGATGCC CTCATGATTT TCCACCTCTT CTTCCTCATTA 360
 AGCATCTTCC TAAGGAAATG AMCATGGCCT GATACTCATT TTGTCACTTG TACAGAGCCC 420
 20 TAAGGATGTT CTGAATTCAG TGGTGCCAAA TAAATGTTGA CATTCCTCTT TTGGTTGATG 480
 GAAGTATCAG TGTGGGAACT GTTGTCTTAA TGGCATTTTA TAAATAAACA AKAKCATATT 540
 AGCAGGGAGG GAGATGATGG AGGAGGGAG AAGTCCATTT GTCTTATTTA TCCTTTTTGT 600
 25 ATTAATAGAG AAGCACTTCA CAGTCACTGG CAATGCCATT TATAGGAAGA AGGTTCTGCA 660
 TTCCTGCTGC TCCCGGAGGG CTTAACTTTT TAATGAAAGA ATAAATGCTC TTCCACTCAG 720
 30 TAGATAAAGT GAAATGTGAA TTGTTAATAA CTGTGCACGG TCAATAAACC GATGTTTTAA 780
 GGAATACAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAACTCG A 831

35

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 590 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

TATATATAGA CNGTTAATAG TCGTGANTGN TGTGNACGAA CATTAACGGA AGTAGCATGT 60
 AGCCAGTCTGA ATACNTATA AGGACAAAGT GGAGTCCACG CGTGCGGCGG TCTAGACTAG 120
 50 TGGATCCCCC GGCTGCAGGA TTGGGCACGA GCTGCCAGGT GAGGAGCAGA GAGACTGTTC 180
 CCTTGGGTGG AGAGGTGTGG GCATGAGAGC CACCCATTGC CAAGCAGCAA GAATGTTCTT 240
 55 GCTTTTTTCC CTTCCTAAAT ATGCAGGGCT CAGGCTCCCA ATTCCGGGCC TGTCTGCTTT 300
 GCTTGTGTTT CTCTGTCCC TGTCTGCCC GAGGCCCCAG GTGGAAGTCA CGACAGGGAG 360
 GGAGACGCTT CCAAAAAACC TGCAGGGCTA TTCCACAGAA TTTGGTTTTT AAGTACAAAA 420
 60

CTTTTGTGTC TGTAAGATAT ATGCAGCCTC ACAGAAGCAG CCTCTGCCCTC CACTTTACCA 480
 CCTACGTTTT TATCTTAAGC ACATGGGGCT CCCTTAGAAC TTA CTCCACT GATTTAAAAA 540
 5 AAAAAAAAAA AAAC TCGAGG GGGGGCCCCG TACCCATTCG CCTTAAAAGT 590

10 (2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1274 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

20 GAGCCACCAC ACCTGGCCTG GAAGGAACCT CTAAAAATCA GTTACGCTT TGTATTTTGT 60
 TCTGTGATGG AGGACACTGG AGAGAGTTCC TATTCCAGTC AATCATGTCG ACTCACTGGA 120
 CTCTGAAAT CCTATTGGTT CTTTATTTT ATTTGAGTTT AGAGTTCCCT TCTGGGTTTG 180
 25 TATTATGTCT GGCAAATGAC CTGGGTATC ACTTTTCCTC CAGGGTTAGA TCATAGATCT 240
 TGGAAACTCC TTAGAGAGCA TTTGCTCCT ACCAAGGATC AGATACTGGA GCCCCACATA 300
 30 ATAGATTTCR TTCACTCTA GCCTACATAG AGCTTTCTGT TGCTGTCTCT TGCCATGCAC 360
 TTGTGCGGTG ATTACACACT TGACAGTACC AGGAGACAAA TGACTTACAG ATCCCCCGAC 420
 ATGCCTCTTC CCCTTGGCAA GTCAGTTGC CCTGATAGTA GCATCTTTCT GTTTCTGATG 480
 35 TACCTTTTTT CTCTTCTTCT TTGCATCAGC CAATTCCCAG AATTTCCTCA GGCAATTTGT 540
 AGAGGACCTT TTTGGGGTCC TATATGAGCC ATGTCTTCAA AGCTTTTAAA CCTCCTTGCT 600
 40 CTCCTACAAT ATTCACTACA TGACCACTGT CATCCTAGAA GGCTTCTGAA AAGAGGGGCA 660
 AGAGCCACTC TGCGCCACAA AGGTGGGGT CCATCTTCTC TCCGAGGTTG TGAAAGTTTT 720
 CAAATTGTAC TAATAGGSTG GGGCCCTGAC TTGGCTGTGG GCTTTGGGAG GGGTAAGCTG 780
 45 CTTTCTAGAT CTCTCCCACT GAGGCATGGA GGTGTTTCTG AATTTTGTCT ACCTCACAGG 840
 GATGTTGTGA GGCTTGAAAA GGTCAAAAAA TGATGGCCCC TTGAGCTCTT TGTAAAGAAAG 900
 50 GTAGATGAAA TATCGGATGT AATCTGAAAA AAAGATAAAA TGTGACTTCC CCTGCTCTGT 960
 GCAGCAGTCG GGCTGGATGC TCTGTGGCCT TTCTTGGGTC CTCATGCCAC CCCACAGCTC 1020
 CCAGGAACCT TGAAGCCAAT CTGGGGGACT TTCAGATGTT TGACAAAGAG GTACCAGGCA 1080
 55 AACTTCCTGC TACACATGCC CTGAATGAAT TGCTAAATTT CAAAGGAAAT GGACCCTGCT 1140
 TTTAAGGATG TACAAAAGTA TGTCTGCATC GATGTCTGTA CTGTAATTTT CTAATTTATC 1200
 60 ACTGTACAAA GAAAACCCCT TGCTATTTAA TTTTGTATTA AAGGAAAATA AAGTTTTTGT 1260

TGTTAAAAAA AAAA

1274

5

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1133 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

15

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AGGATTTTTC	CTTGTTCAAC	CAAAATCTGA	GCATTCTTTC	TATGTTGAAA	ACACTGAAAA	60
ACTAATTTWA	GTTAATGAAC	TAGAAAGAAT	ATTGATTTTW	AAGAAACAGA	AAAATACTAC	120
TTATTTTCCT	TCTCAAATAA	CGTTTCTTTC	AAAACTTCT	GGCTGAAGTA	TAACATGCTG	180
GTAGTTAACA	TAAATCTTGT	CTTCTCTTTC	TTCTTTATCT	TTCTTTGTTA	TTTACATGCT	240
TGTATAAATG	TCCTTTGTTT	TTATTAAGTG	CCTAATTGAC	AGAGCTTAAT	TTGAAGAAGT	300
GCCCTAATTT	ATTGACCACT	TAAGAATTGC	CTTTATTGGG	GTATTTTATT	TGTTCTGCG	360
TCCTTTTGAT	GTGTTTCAGT	CTACTCATCC	CTGTGAGTAT	GTGTGGGGGA	CAGCTGATAG	420
AAGGGAGGAG	AGTGTGTCTA	TGCTCAGGAT	TGCCCTTTAG	CCACTCAGCC	AGAGATCCAC	480
AGGGAGCAAC	AAGGACAGTT	TCACATGCTT	AGACTTTCTT	GGAAGAAACA	GTGAGGAGGA	540
GTAAGTCGTG	AGTAGTGTCA	AGCTGGATGT	AGAATTGTCC	TAAGGCAGTT	GACCCACCT	600
TCCAACATGT	TTTCACTTTA	TTTCCCCCTC	CCTACATTTG	GGTTAGGTTT	CATTTGGATT	660
TGCAGCAATA	ATGACTTTAT	TTCTCTCTTG	GTCAGGATTT	GGCACATAAA	ATCCTTTTAT	720
TATAGAACTA	GCTATTTTAG	TTACATAGTA	ATGTAAGTAA	TGGAGAGATT	TATAGAGAAT	780
TTTGKTTTTG	CTGTCAATAA	TOTCCATTTT	GGAGACAGAT	ATGATAGAAC	TAGAAATTAA	840
GTTGCATTTT	TGCAAGTGCC	ATTTGAATGA	ACTTCAAGTA	TCTTCTTAAT	TATTAAATTT	900
TCTGATGAAG	GCATTGTAAAC	AAATATATAG	TATTATTAAA	TCTAATTAAT	ATTTGGAAAT	960
ATTAATAAAT	AGGTATTTTA	TTTACTGTAA	AAAGTCAAAC	TTTATTATGT	AGATAAATCT	1020
TATTCCTTTT	ATTCTTTCCC	CTGTTTACAT	CCTTTTACAA	AAGCTTAGTC	ACCAATTAAA	1080
GCTTTCCTAT	CAAAAAAAA	AAAAAAAAAA	ACTCGAGACT	AGTTCTCTCT	CCT	1133

55

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

60

334

- (A) LENGTH: 661 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

CAATTCGGCA CGAGGGGAAA AGGATGCTGA ACGAGAGCAG AAAGCCTCTT TCCTTTGCTT 60
 10 CACGCCCTTC CAGTCTTTAT TTAAACTCG GGTTCCTTTT CTGTGGTGGC AGCAACCTTT 120
 ACTCCACCTG CACTGCTGCT COTGGGGGCT CCCCAGGCCT CCTCTGCTT TTCTACCCAG 180
 TGGCTGACGG GATGCTGTG TTGCCTGGAC GCACCACTGC TCTCTGTCC CTCACCTTGG 240
 15 CTTTTGCTGT GCCCTGCTCT GGGGTTGAAG CTGGCCCATG TGTCCCGGG AGTCATGGCT 300
 GCTCCTCCTG GGAGGCTCT GTGTGCTCA CTTCTCCAC ACCTGGGGCC AGCTGGGGAG 360
 20 CCGGTGCTCT GTTCCCTCG GCTGCTTGGC ACAGAGTGC AGCCTGGGAY TCTCCGTGA 420
 CCCAGACTGG GCATTTTGGC AGGGGGGCGA TGGGAGGAGC AGGTGCTTTG CCGGGGGCT 480
 GTGTCTGCAT TTCTGGACGC CCCAGAGCAC AGAAGTTGCC GGCACCTTGA GGTCTTCTC 540
 25 GGCATGTGCC AGATTACATG AGTGACGGCT GGGAATATGT TTTCTTTTGT GTAATGGAGG 600
 CGTGTTCAC ATATACTAAA GCTCACCAA AAGTAAAAA AAAAAAATA AAAAACTCG 660
 30 A 661

35 (2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

45 ATTGGGTACC GGGCCCCCCC TCGAAGTTTT TTTTTTTTTT TTTTAATGAA AGCTCTCAAA 60
 TAAGCGATTT TATTCCTATC CATGATTGCA GACATTTACA AAACCATAAC ATCTGAGTTC 120
 ACCTTAAAAA ATAACCTATA TAAAGCAGTG ATATACACAG CACAAAATAG TTCAGGGAGG 180
 50 GGGCAGGAGC AACTTGTAAT AATTAAAATG TAAACGTGAA AAAAAGGATG GAATAAAAGT 240
 CCCTACTTAT TTCTACTTAA GATGTCATGT GATAATATTT TACAATGTCC TGTGGGTCAA 300
 55 TGTATGTATG TGTATATGTC TGTATAACAT ACACATATAC AGTACATTTCT CTTTCCACAC 360
 CATATACATA CACACATAAT TATTTGCACT TCAGTTTAGG GCAATTTCTAA TATGCCACTC 420
 CGTACAGTTG TTTGAATCAC AATTGGACCC GCTTTCTTCA CAAAAGAGGG GAGAGAGCAG 480
 60

5 GAAATAAAAA GCTTGGTTTG GTGTGACTGA GATTCCCTTTG TTAACTGTA CACTGTGATG 540
 AATAATTTTC TTCCGTAGTA GTTCTGTGAA GGGCTGACTC ACTGTGGPTT TCATGAGGAG 600
 10 ACTTGGTAAT GGATCACACC CTCATTGTCA TGCTAGGGGA GTAACTCTCA CTCTGAAAAG 660
 GATTTAAGAA ATTTCCCCC ATTTGCCCAT CATCCCTGG AGTGCCCGGT TCATTACTCA 720
 GGCTCATATT ATTGGGAGAA TTCTTGGAAA TACTGTCCAT ATCTCCTGAG CCTAAAGAGC 780
 CATTCATGTG ATGTGACTCC ATTCCTCCTA ATCCACCCAT GGGACCATCT GACCCAGGRC 840
 CCATTGGAAA ATTAGGTCTG TTAGGTCCAG GAGGTACTGC ATTCATTAAA GTATACATGT 900
 15 TATCACCAGA GTTGGTTGAA TCTGTGGAC TAGGCATGAT GGGTGTTCCT GGTGGCCCTC 960
 CACCTCCTGG AGGACCTACA TAATTCCCAG GAGATGCTGA GGAGTATGGT ATTGAATTGG 1020
 CATTTGTTGG GTTTGGCCAA GGTCTACCAC CACCTGGACC CATGTTTATT CCAGGCATTC 1080
 20 CAGGGCCACC TAAAGCATTC AGTGGGGGTC TCATTGCACC TCCATAGTTC TGTGGTCTTA 1140
 AGGGCAGCAT TCCTCTTGA GAGTCATTC TCTGCATTC CCCACCCATA TTGGATGTC 1200
 25 CTTGTTGTG AGTTGGATCC ATTCCACTGG GGAGTAATGG CTGACTTCCT GGGACACCTC 1260
 CAAGTGCTG ATTAGGTATC CTCAATGGGG GCCTTGGACC TCCAGGGTAC CGAGGTGACA 1320
 30 TAAAGGGTA ATCATGGAAG GCTTTTGCTT CACTTGAGTG TTCACATGTT TCACGTCT 1373

(2) INFORMATION FOR SEQ ID NO: 81:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

45 ACTTTGTCCA AATGTGTCTG TCACATGTAG TCAGCTGNAG NAATTTAAAA TGAATTGCCA 60
 AGTGAAGAGT CTGTGGATTA ATTGGCCGTT AATTAACAGG CTTTATCAAT GTGTCTCAA 120
 GGGAGAGGCC CAACCCTAAT TAAGGAGCTA AACTTCCTGA GTGAGGGCCT GTGAGGATGG 180
 50 AGGTGGAGGA GGCATCTGGG GCGGGTGGTG GCGGGGCCAG CAGATGGCGC CTCCTGGCT 240
 GAGCTGCCCC CACCGCCAGT TCCCTCATTT CCACTCAGGA AGGCAGAGAA GGCAGAGTGA 300
 TCTCCTCAAG GAAGAGCTTC CCCAGCCTTC GGGAGCAGCT GGCAGGGCGT CCGGAATAA 360
 55 GGCCTACAG CCGCCGCTG CTTCCAACTC ACTAACCTG GGCCTCTGT CTTTCAGATT 420
 CAACCGGTTT AACAGAAGCC ATCCCCAGCC CAGCTTAAAT TATAAAGATA GACAATAACT 480
 60 CTGTTCCAAT CTGCGTGGTG CTTCTTTAGT AAATACTGTA CAGATTTTAC CATGGAGAAC 540

TTTTITTTTA GTTITTTACCT TTICTTAATT ACCCTTATTC CGAATGGAGG AACACTTTCT 600
 ACCACTGCTG ACCATTGTAA AATACCGTGT ATATAAATCC CATTGRAATA ATGCCCTGGA 660
 5 ATAGAACATC TCAATGCTG CTTAATTACA GACTCAGGTC GATTACTTGT ATTTCAATGA 720
 ATGTTCTCTC AAGTTAGACA TCTGGTGCAA GACCAACCGG GAGACCATGG AATTGTCAAA 780
 10 AGTACAAACT GACAGTGTGT ATATTTAATT TAAAGACTTA TTTAAAAACT CACAAGCTCT 840
 CACCTAGACT TTGGAGAGCA GTCTGTTTTC TGTAAATGTCT GATACTAGAA ACTAATTGTC 900
 TTATTTTACT TGTATTCAAG ATTTGAAGAT GTATTTTATA GACAAGTTCT GTTTTTGAAC 960
 15 TTTGTGGAAC TGTTCCAATC AATCAATTTT CCAGTTATGA TGAGTATTTA CATTATGAAT 1020
 GTATAACCCA GACATGATTT GTAAAGCCGA CAGTATGTTT CTATTACACA ACACTTTTTG 1080
 20 ATACAGCGTC TCTTGTCTTC ACTGATACTG GAGTCTCCGT TGTCTGCMNG GTCCCTTCGA 1140
 GTTCTAGATT ACAGACACAA TCATCTGTG ATTTTATTTT TAATATGGAT ATGCTATCAA 1200
 ACTGTGATAC ACTTATAATT CACTGGTCCT GCATCAGGAG ATGGAGTGGG GAAAACTGTA 1260
 25 TTTAATAEAG TTTGTATCTG AATAATCTGT ATGGTTTATA CAGTTTGTGT TGTTCAGAGA 1320
 TGTTTAAGT TGTATCTTTG TTTTCTAAA GATTAAAAAA GCACCTGCCC CACTGTAAAT 1380
 30 ATACAGCATG TAAATTTCT RTAGTATATA AATGGCAGCA AATCACAATA AAAAAAAN 1440

35 (2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1381 base pairs
 (B) TYPE: nucleic acid
 40 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

45 CCCCCGCTGC AGGAATTGK YACGAGGCCA GCAGTTGCTC CCAGTTCAGG AGGTGCTCCT 60
 GTACCCCTGGC CACAGCCCAA TCCTGCCACT GGTGACATCT GGGGAGACTT TACCAATCT 120
 ACAGGATCAA CTTCCAGCCA GACCCAGCCA GGCACAGGCT GGGTCCAGTT CTGACCTGAG 180
 50 CACGGTTTTT CCTCATGTGA CTTCTGGGAA GCGCTCCCT CATCTGGGCC AAAGGAAGGA 240
 GGACGAAGCC CTCCTCAGCT GGCCTGTGTT TGGGGCATGA ATCTCTCCTC TCTCCTTGT 300
 55 CTGGCTCTGT TGACAAACCG GGCATGTTTG GCAGTAAAT GGCACCGTGT CACACTGTTT 360
 CCTGGGATTC AAGTATGCAA CCAGAACACA GGAGAAGAAA AGCTCCAGGA TCCCTGTCCC 420
 CATCTGTCTT CTTGATGTGA GAGAGACTCT GAGACTTCTT CCATCGCAAT GACCTGTATT 480
 60

337

AAACACAACC CCCCCAAGCA AAAGAAGAGG TTGAGTTTGC TGCCAGGATT CAGATCAGCC 540
 CTTCGCCAGG TGTGCAGGTG TCACATGATC ACAGTTTCAGC GGGAGGCTTT CCGTACCCAC 600
 5 ACTCGCTGTA GCACTTCAGT CCATCTGCCC TCCAGAGGAG GGTTCCTTCC TGATTTTATG 660
 CAGGTTTAGA GGCTGCAGCT TGAGCTACAA TCAGGAGGGA AATTGGAAGG ATTAGCAGCT 720
 10 TTTAAATATG TTTAAATATT TTGCTTTGCT AATGTGCTGA TCCGCACTAA CTCATCTTTG 780
 CAAAAGGAAC TGCTCCCTCG GCGTGCCCCA GCTGGGGCCT CTGAAGGGAT TCCTCACTGT 840
 GGGCAGCTGC CCTGAGCTTC AGGCAGCAGT GTTCATCTCT GGGCAGTTGT CTGGTTTCCA 900
 15 TGTATTCTAG GCCAGGTAGG CAACACAGAG CCAAGGCGGG TGCTGGAAGC CAGACGGAAC 960
 AGTGTTCGGG CAGGAAGGTG GATGCTGTTG TCATGAGCT GTGGGAGTTG GCACTCTGTC 1020
 TGCTGGTGGC CCTCTCGGCT CACATGTTCA CAGTGCAGCT CCTGGCAGAC TTGGGTTTTC 1080
 20 TCTTTGGTGG TTTCTAAAGT GCCTTATCTG CAAACAACCT CTTTCTGCT TCAGGAAC TG 1140
 TGAATGGCTA GAAGAAGGAG CTCAGTAAAC TAGAAGTCCA GGGTTGCTTG GTTTACTGGT 1200
 25 TTATAAGAAA TCTGAAAGCA CCTCTGACAT TCCTTTTATT AACTCAGCTC TCAGTTGAAA 1260
 GATTCTTTCT TTGAAAGGTC AAGACCGTGA ACTGAAAAAA GTGTTGGCCT TTTTGCGGGA 1320
 CCAGATTTTT AAGATAAAAT AAATATTTTT ACTTCTGTCA AAAAAAAAAA AAAAAAATNT 1380
 30 C 1381

35

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 1706 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

45

ACTGCACCAC TGCCCGAGTC TCCCGGCTGG ATGAAGACCT GGTCCATGAG GAAGCTGGCT 60
 AGCTCAGACT GGAGAGTAGC TTCAGGAAAA AAGACAAGTG GCCTAAGGAA ATCAGCGGCC 120
 50 CCAACTATCA TCTGAGGGCT AAAGATGACA AGTAGATCAC TTAAATAGAC AAAAGCCTGT 180
 -----AGGGGGAAAA GAAAGGATGT TAAAAGGAC AGAATGTTTC CCAAGGTAGA AATGACACTG ----- 240-----
 TCAATTTCTC CTTGGAATGG GGGCAGGCAT ACTCGCCTTG TTGCTCCAC TTGAGTCAGT 300
 55 ACTCAGCTGC TCCTGGATCT CAGTATCCAC ATCTGAGAGG CAACTCTGGC AGACTTCACA 360
 GAAGGCCACC ATTCTGTCCC TCAAACGGA CAGCTGCTTC TGTGGGCACA GTGGCTTGAA 420
 60 GGGGAAGAAT GAAGACACAG ACTCCTCTGT TCCATTATC CCATCTAAGA CCCACACTCA 480

5 CCTGGGGAAG CATCTGATTT AGAAATGTGG GTTAGTGTCC ACAGAATGGA AAAATAGACA 540
 AGAGTCRAGG CTGCCAGGAT AACCTGTAAC AACAAACGGT TTGAAAAATG AGGTTTGGGT 600
 TAGGAGAGGG AGAGACAGAT AGCCAGAAAC ACACCACTGA AGAGGAGAGA AAATGAGTAA 660
 ACGGAGAGCT AATTCCTTTT CCAGTGGAAA ATGAGTGATA TTCTGGACAT TCTTCAGAGG 720
 10 CATCTACACG AAGTAGAAAT GTCACCGCTC CCAATTTTAC TCTACGTCTT CTAGAATCCC 780
 TCAATATTAT CCTTGGCTTC CAGGAAATCC AAGAAGACCC TGGAAAGTAGA GTCCACCTTC 840
 15 TAAGAGAGGA ATGTAAGAGG TGACCCCCAC CCACCTGATC TTCTCGCTT TGTCCACTCC 900
 ACGCACTGAG ACTTGACACA CCTAGTGGCC ACCTAGAACG TAGGTCCTTA AAATYTAGCC 960
 CCCCAGCCCC CAACCCATCT CTAGCCTGTC CACTCACCTG GTGAGGAAC/ TYTCTGTGT 1020
 20 CCACAGCTTT CTGCAGGAGT TGGCAACATG GTCATAGAG CTCCCAGCGA GTCAGGTCAT 1080
 GAGTGCTTTG GCGGAGAAAG GGAATGTTA TACTGGAAA GAACAGAGGG AACCAACTCC 1140
 25 ACAGACACCA GTAAAAACGG GATGGGGAAG AGGAGGAAAG CCACTCAGTT GTAGAAGGCA 1200
 GAGAGGCGTT TCAGAGTGGC TGCCAGATTA TATACCTCAT CCTCATCTAG GAAGGACGAC 1260
 TGAGAAGGAA AGAAGATCCA CAATAGCATT TCCCCAGAA CTCATCAGTC CACATCCCCC 1320
 30 GTCTTCAGC CCCTCCCACC CTTGTTTGGG GTGTCCCAT GTCCAGCCCC AGCTCCTACC 1380
 TGTAACAGCT CTTCAAGCTC CTGCTGGAAR CGGTCAGTCA GCAAATCTAC TAGCTGGCTG 1440
 35 CGGGCAAAGT CCGCCCCGGT GAAGAAAGTG AATTCGGGAT TACAGAGCAG GTAAGAGCAT 1500
 GCGCCCCAGC CTCAGGCACC GCTGGCTCTG CATGCTTCAC CACCACCTCC TGGAGTTGCT 1560
 GCAGGAACAG CTCACGTGC TGAGAAGAAA AGGCAGAAGA TGGTGTGCTG TGGGGATGGG 1620
 40 AGGAGGACAC TCTTCTGGCG GGAAGTGAA CGGGTTTAA AGCATTAAAC TTCAAGGATA 1680
 AGATGCCTAA RAAAAA AAAA 1706

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS;

- (A) LENGTH: 573 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

GAATTGGGCA CGAGCTTGGT AGCCTTAGAA CTGCATGACC TGCTTTACCA CTGGGAAACA 60
 CGAGCACAGC CTAGCTTGAT TTGTATGTG GTATCAGATC TAAGGTGGAT GGAATTCAGG 120

ACTTCCTGTC TACTCTTTGA TTTTGTTTTA TTTTGTAGAA TGTMTTATTT TGTMTTATTC 180
 ATTTATTTCAT CTTTCAGAGAC ATGGTCTGGC TCTGTTGCCC AGGATGGACT GCATGGTGTG 240
 5 ATCATAGGCC ACTGCAGTGT TGAGCTCCCG GGCTCAGGCG ATCCTCTCTG CTCAGCTTCC 300
 TTAGTAGCTG GCACTATAGG CACATGCCCT ACCATGCCCTG GCTTGTCTA CTTTTTGAAT 360
 10 GATGTC/CAA ACTAGAAGGT CTATTAATTT AAAAAATTAA GGATAGCATG CCATAATTAA 420
 AAATAATAAC AGTGGGAAA GGCACCTTCC AATGATTCAG ACATCAACTT GTGATTAA 480
 AAAACGAAAA ATAAATAATA GGAaaaaaag GGGAAAAAGT TAAATAAAAA TAAAAATTAA 540
 15 AAAAAAAAAA AAAAACTCGA GGGGGGCCCC GTA 573

20 (2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CTCTTTGGCT GTGTCTACCT CTTTCATCTG CTGCCCGGAC ATAAGCACCG CCCTGCCCCCT 60
 AGGCTCCAGC CGTCCCGCAC CAGCCCCCAG GCACCGAGAG CAGGAGCATG GGCACCAAGC 120
 CAGGCTCCCC AGGCTGCTCT YCAGCTCCCT TATGCCACTA TCAACACCAG CTGCT/GCCCCA 180
 35 GCTACTTTGG ACACAGTCA CCCCCATGGG GGGCGTCTCT GGTGGGGGTC ACTCCCCACC 240
 CACGCTGCAC ACCGGCCCCA GGGCCCTGCC GCCTGGGCCT CCACACCCAT CCCTGCACGT 300
 40 GGCAGCTTTG TGTCTGTGA GAATGGACTC TACGCTCAGG CAGGGGAGAR GCCTCCTCAC 360
 ACTGGTCCCG GCCTCACTCT TTCCCTGAC CCTCGGGGGC CCAGGGCCAT GGAAGGACCC 420
 TTAGGAGTTC GATGAGAGAG ACCATGAGGC CACTGGGCTT TCCCCCTCCC AGGCTCTCTG 480
 45 GGTGTCATCC CTTACTTTTA ATTCTTGGGC CTCCAATAAG TGTCCCATAG GTGTCTGGCC 540
 AGGCCACCT GCTGGCGATG TGTCTGTGT GCGTGTGTGG GCACAGGTGT GAGTGTGTGA 600
 50 GTGACAGTTA CCCCATTCA GTCATTTCTT GTCGCACTA AGTCAGCAAC ACAGTTTCTC 660
 TGAaaaaaaaa aaaaaaaaaa AAAC 684

55

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1036 base pairs

60

340

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TGGAGGCAGA TGCACAGGAG AAAGTTCCCT GTCCGCACCC TCTCAGACCT GAGGCTGACC 60
TTGCACTGAG GCCTTCTCCT CGGCCCTCG CCGCCCCCA GAGCTGCCAT CCCTGCTGTT 120
10 ACAAGCCAGA GCAGCCCGGA TGTGAGGCC CAGATCACCT CCAGGACTT GGGTTCCCA
TCTGAAATCC TTTATTTTGT TACCATGGGG TGGCCCCGG GCTGAGAAGG AAGAAGCACC 240
15 CTCTCCCCGG CCTCCTCTGT CTGCACCCGT GGGGCTGTGA CTTACTCCTG CCTCCAGGGG
CGGGCCGGGG CCCCCTGGGA CCTCTTAAGG CCCAAGGTGG GCCCCAGGAC CTYTGCCAG 360
AGTGGAYTGC TCATGGCAGA TGTGTGGCAA TGTCTGGCTG WGTCTTTCCG GCAMCTGCCT 420
20 YCCCTYTCCC GGGYTCCCT GCTGCATGGT GGATGTGCTC CTTCTGGCC CCGTCACATT
GCCTCCTTGA GCCTTAGTCC AGGGGTCAC TYCTCCCACC CCACCTACCT CACAGGGTTG 540
25 TTGTGAGCGT GCACAGAGGA GCAAAGTCCC TGAAGGCCCT CAGGCAGTAT ATAGGGGCCG
CCCACCTTCA GCTGCCCTGG GATGGGAAGG ACCCAGCCCG ACCCCTGGGC ATAACACTGT 660
GTTTGCAAAT GCAGATTGAG GTATTGGGA TGCAGTTGT GGGGAGCTGG CCTGGCAGAG 720
30 TAGGGGTAGT TGGCTTGCC TTCTCTTTGG TGATCCCACC CCCAGCCATT TGCATTGCTG
GCCCAGCCCG TGGCCTGGGG GCGGGGAGA GGCAGCAGAA GGGGCTGGC AGGGGCGGTG 840
35 GAGGACTCAG GAACTGCCCG GGGAGAGTGG GTATGGCGGC TGAGCCAGGG GGCCTCCTGT
GTTTGACTTC CCGGGATGGG TCCTTGCTTC TCAGCTGTGT CCGACCCAC CATGTAATAA 960
AACCCAAAGG AACAGCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1020
40 CCCNGGGGGG GNCCCG 1036

45

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

50

(A) LENGTH: 908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

TTAAACAAAT GGAATCATGC AATATGTGAC CTTTTCGCTC TGGCTTATTT TATTTAGCAT 60
AATGTTTTTG AGGTTTATCC AAGCTGTAGC ATGTATCAGC ACCTCATTTT TTTTCTGGC 120
60 TGAATATTAT TCCATTATAT GCAATTACCA CAATTCAATT ACCTATTCAT CTTTGTGTTT 180

	TGCTGTCTGG CTATTGTGAA TAATGCTTCG ATAAACATTC ATATACAGT TTCTATGTGG	240
5	CTTTATGTTT TCATTCTCT TGGCTATCTA CATGGGAGTA GAATTCCTAGG TCATAATATA	300
	ATTTTATGTT TAACTTCTCA AAGAATTGCC AAAAGGTTTT TCATAGTGGC TGCATCATTT	360
	ACATTCACCAC CGGCAATGTA CAAGGATTTT TATTTTTCOA TATCCTTGCA CTTACCAACA	420
10	CTTCPTTTTK GTWATWATTT TGTPTTTTCA TTATTGCCAC CCTAGTGGAT GTGAAATGCC	480
	ATCTTATGTT TTGATTTGC ATTTCTCTAA TGACAAATGA TATCATACTT TTTTATGTG	540
15	CTTACGGATC AAAGGTATTT CCTTGGAGAA ATGTCCCTTC AAGTCCTTTG CCATTTCAAA	600
	ATTTGGTTAT TTGTCTTTTA TTATTCAGTT TTAAGAAATT CTGGCCAGGC GCAGTGGCTC	660
	ACCTGTAATC MTAGCACTTT GGGAGGCCAA GCGGGGCAGA TCACTTGAGK TCAGGACTTC	720
20	GAGACCAGCC TGGCCACAT GGTGAAACCC CATCTTACTA AAAATACAA AATTAGCTGG	780
	GCGTGGTGGC AGGTGCATGT AATCTATCT ACTCAGGAGG CTGAGGCAGG AGAATCGCTT	840
25	GAACCCAGGA GCGGAGGCT GCAGTGAGCC AAGATCACCC CATTGCACTC TAGCCTGGGT	900
	GACACAGA	908

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 555 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

40	TGCACTGGTT CCTCTCCCC AGCAAATACT GCCTTCTGT TTTCTCTGA TGTGGCAGGT	60
	GACTACAAAA TCCGCCTTGG TATTCTTCAA ATGCATATAT ATTCTTTCT TGTGAGCTCC	120
45	CTCTCTCCT AGATTAGAAA ACTGCCTCAT TTTCTGTCA CTGGATGTGC AGTCCCAGCT	180
	TGTCTTCCTC TCCTCCCCCC CTGTTGCAGG TGTCTTTTT TTTTCTCTC TCTCCCCACT	240
50	GGGCAGCAA AGTTGTTCCA CAGTGGAAAW TTAGGCATCC TCAAGTTTCY TCCCAGCTTC	300
	TGCTGTGTTT TCTTAGAGTA AATTGCCAAT TTCTGTTTT ACAGGAAATC CTTTTTTAAA	360
	AATGGAATCA GTGTGGTCCC CATCTACTCT GCAAAAATTG CATTTTTCTC TATTTTCAA	420
55	TGAGATTTGT TCAAGTTTCA AAACCACGTG AAATAATAAA TGTATAGTAG TTTTCTTTTC	480
	CTTGGGCATT GCTWGATATG TGAAATGGGT TTATGAAAA TAATAAATC ATAACGCTAT	540
60	TTGTTTGACT TTCAATTTCA TGGGAATTTT TCTCAGCTAA ACTCTAAATG GTGATTARGC	600

AAAAAAAAA AAAAAAAAC/ GRAGGGGGGG CCGGTACCAA TTCGCCCTAT AATGA

655

5

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1102 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

15

TTTTTTTTTT ACCATTAAA ATAAAATGAA AGTGACCTTC TGTTTATAAA AATCTTTGTC 60

TGCATCTCTG CTTATTTCCT TAGAAGAGAT TCCAAGAAGC GGTGAGTGAT TTCACGGCAG 120

20

CAGACGGTTG GGACATATTA CGGGCGCGGA TCCCTCTTGG AGTGACATGA CTCTCCGGAG 180

AGATTTAGTC GTCACCTCTG CCGTGTGAGG TCGGTACAC CCCAGGGATG TGTCTATCAA 240

25

GATGGAAGAT CTTTTACACG CTCTTGATTT TGTTCGCT/ TTTTCTATT ACTAGTGAGA 300

AKGAAACTTT TTATATGATT ATTATCCATC ATAATCCAAC ACAAATTACT GCTTCATGTT 360

CTTTTACTTT CCGTGAAGG TTTTAGTGCC TTTTAAAAAT TGCTATATAT TAAGCTTGTT 420

30

AATACTTCCA TGCTGTATTT GTGGSCATCA RTTCCCCCG GNACAGGCT GCACATTTTG 480

CCTTCACACG CTGGGTGGTT TTTCAATTTT AMTCTATTT CTCGTCTTC TATCGTTTFA 540

35

TGTTACAGAC GGTTCCTCCG TGTAGAAAGC AGTTTATGAA GATTTACTTT CGACAGTCTT 600

CTCTCTACTT TCTACAGTGA ATTCTCTGAT GTGTCTGGGA GTTTGGGGGT CTGGGTAAGA 660

RTCTCTCTCT CACCCTATTC TCTATTACGA TCCACAGCCT CATGCTTTAT GARATTGGTG 720

40

GCGGGGARGG GGGGAGATTT GCGGATCCCC CAAGCCAGAC TTTATCCCCC TATCCCTGCC 780

TCTGGATCCC ACCTACAGGC CTGGGAATTC CCGTGGGTA GGGGCGAATG GTCTCGCACT 840

45

CTCACCTGTA CCCCAGGGCT GGCACAGGAT GGTCAAGGAG AGAGGCTGCC CAAGCGCATC 900

CYTCTGGTGT CCCCCTGACA CGCCTCCAAA GTGAGCAGGT AGTTTCAAC AGCCCCACGT 960

TGCAGGTGGG AGATGAAGCT CAGGCTGGAG ACCAGTATCT CACAGTTCTC TTTGCATGGC 1020

50

CGGGTACTTG TTAGTCAACT GATCAAGTGA AAATCTAGC CCCAGAGGCA GGAGAATCCG 1080

GAACAAAATT AAACCAGCCA GG 1102

55

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1533 base pairs

60

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

	GGCACGAGCC GNCACGGGCA GCGCCCCATA GCGCCAGGGA CCCCCCTGGCA GCGGGAGCCC	60
	CGCGTCGAGG TTATGGATCC AGCGGGGGCG CCGCGGGGCG TGCTCGGGCG GCCCTGCGCG	120
10	TGCTGGTGC TGCTGAACCC GCGCGGGGCG AAGGGCAAGG CCTTGCAGCT CTTCCGGAGT	180
	CACGTGCAGC CCGTTTGGC TGAGGCTGAA ATCTCCTTCA CGCTGATGCT CACTGAGCGG	240
15	CGGAACCACG CCGCGGACCT GGTGGGCTCG GAGGAGCTGG GCGGCTGGTA CGCTCTGGTG	300
	GTCATGTYTG GAGACGGGCT GATGCACGAG GTGGTGAACG GCCTTCATGG AGCGGCCTGA	360
	CTGGGAGACC GCCATCCAGA AGCCCCCTGT TAGCCTCCCA GCAGGCTCTG GCAACGCSCT	420
20	GGCAGCTTCC TTRAACCATT ATGCTGGCTA TTAGCAGGTC ACCAATGAAG ACCTCCTGAC	480
	CACTGCACG CTATTGCTGT GCGCGCGGCT GCTGTCACCC ATGAACCTGC TGTCTCTGCA	540
25	CACGGCTTCG GGGCTGGCC TCTTCTCTGT GCTCAGCCTG GCCTGGGGCT TCATTGCTGA	600
	TGTGGACCTA GAGAGTGAGA AGTATCGGCG TCTGGGGGAG ATCGCGTTCA CTCTGGGCAC	660
	CTTCTGCGT CTGGCAGCCC TGCGCACCTA CCGCGGCGCA CTGGCCTACC TCCCTGTAGG	720
30	AAGAGTGGGT TCCAAGACAC CTGCTTCCCC CGTGTGTGTC CAGCAGGCCC CGGTAGATGC	780
	ACACCTTGTG CCACTGGAGG AGCCAGTGCC CTCTCACTGG ACAGTGGTGC CCGACGAGGA	840
35	CTTGTGCTA CTCCTGGCAC TGCTGCACTC GCACCTGCGC AGTGAGATGT TTGCTGCACC	900
	CATGGGCGCG TGTGCAGCTG GGTTCATGCA TGTGTTCTAC GTGCGGGGCG GAGTGTCTCG	960
	TGCCATGCTG CTGCGCTCTT TCTGGCCAT GGAGAAGGGT AGGCATATGG AGTATGAATG	1020
40	CCCTACTTGT GTATATGTGC CCGTGGTCCG CTTCCGCTTG GAGCCCAAGG ATGGGAAGG	1080
	TGTGTTTGCA GTGGATGGG AATTGATGGT TAGCGAGGCC GTGCAGGGCC AGGTGCACCC	1140
45	AAACTACTTC TGGATGGTCA GCGGTTGCGT GGAGCCCCCG CCCAGCTGGA AGCCCCAGCA	1200
	GATGCCACCG CCAGAAGAGC CTTATGACC CCTGGGCGCG GCTGTGCCTT AGTGTCTACT	1260
	TGCAGGACCC TTCCTCCTTC CTTAGGGCTG CAGGGCCTGT CCACAGCTCC TGTGGGGGTG	1320
50	GAGGAGACTC CTCTGGAGAA GGTGAGAAG GTGGAGGCTA TGCTTTGGGG GGACAGGCCA	1380
	GAATGAAGTC CTGGGTGAG AGCCAGCTG GCTGGGCCCC GCTGCCTATG TAAGGCCTTC	1440
55	TAGTTTGTTC TGAGACCCCC ACCCCACGAA CCAAATCCAA ATAAAGTGAC ATTCCCAAAA	1500
	AAAAAAAAAA AAAAAAAAAA ANCCCGNGGG GGG	1533

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

ATCCTCTGGA ATCTAGGTGG AAGCCACCAA GCCTTCTTCA CACTTGGGTT CTGAGCATCT 60
GCAGACTTAA CCCCATGTGG CAATCACCAA GGCTTATGGC TTGTGTCTTC CAGAACTGTG 120
GCCAGAGCTG TACCTGGGCC CTTTGTAGCT GAGGCTGAAG CCAGAGTCTG AAGCTCAGCA 180
GGGCAGTARG GGCCTGGGCC TGGCCCTGA AACCATTCTT TTCTCCTAAG CCTCTGGGCC 240
TTTGATGGGA RGGCTGTGCC TCAAGATTTT TGAATGCCT TTGGAGGTT TTTGCCCTGT 300
CTTGATATT GCCTTCTTT TAGTTATGCT CATCTCTCTA GCAAGTGAAT GTTTCACAC 360
CTGCTTGGAT TCTTTCTCTA CCACAGARCC AGGCTGCAAA TTTTACAAAC TTTTCACTC 420
TGTTTCCCTT TAAATATAA ATTCAATGT TAAGTCACTT CTTTGGCTCC ATATCTGATT 480
TAGGTTGCTG GAAGTAGCCA AGTCACCTCT TGAATGCTTT GCTGCTTAGA AATTTCCTCT 540
ACTAGGTAGC CTGGGTCATC AACTTAAST TCAA 575

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 639 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

TCCTTTCATC TTAAGCACCA CCGACAGGG CAGGTACTAT TACCATCTCC GTTTGACAGA 60
TNAGGAACCT GGCACAGGAA GCATTTAAGT GGATTCCCA GCATCGCCCC ACTGTCAGGA 120
GCAGANTCAG AATGGGCTC AGCATCAGGC TCCCAATCCT GGCTTCTAAC TGCTGCGCTC 180
TGCCCTTCYC TCWCCCCACC TCCCACTCC AGTGCCCTTG GTCATGCCAC TGCAGCTTTC 240
AGGCCAATAC TGGATTAGCC TCTTAGTGT CTTGTCCCTG CAGCCATTTC CCCAGGCAGC 300
AATTCCATGT GGCCTCACTG ATGTAGGTGG CTCTGTGTG ATTTGTGACA TCTATTGAA 360
TTGTTTATGC ATCTTGTTC CACTCACAGC ACCCTCCCTC TCACACGTC TCTTATAAA 420
AATGTCCCTC AGTGTCTGCT ATGAGCCAGG TGCAGACTTA AGTGACAGGG CTGCTACGGG 480

345

AAATAAAAA TTRACRAGGA GCRCTGCOCT CMTAATGCAC AGTRACAAC TATGTTAAGT 540
GTCAGGAAGG AAAGGTTAAG GATGCCAGGA AGGCTTTTAA TAAATAACCT GACTTAGATG 600
5 GGCAGGTGCT GCTGARGATT ANGAACGTGT TCTTCTCGA 639

10 (2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 744 base pairs.
(B) TYPE: nucleic acid
15 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

20 GAATTCGGCA CGAGAGTGGC TGGAGTCTGG CTGCAGAGGG AAGACATCAG CAGGGAGGGA 60
GCCAGGGCCT GTCACATCTT TCCTCTGGCC ATTGTCCTGG TCTTTGTAAG CCCAGAATCT 120
25 CCCCTTCCTT GAAGGGAGGC CAGCACCCCA GGAGGGCAGC AGGTGTGCTG TGAGGGTTGG 180
AGTAGTGTGA GAGGTCAGCG TACACTAGAA/TGGCCATGGA CACCATGTGG GGGTGCTCTG 240
GGCTGGGCCA CAGAACAGTG TCCTTCCTGC TGCTCCTCCC CTGCAGCTTC CCCCAGCTTT 300
30 GTNGTTTATT TGGTTTGATA CCAATCAGCA GACCCTGCAA GGTGGAAGCT CCCAGGCTCT 360
CAGTCCCACS ACTCTCATGT GCCAGTCACC CMTACTGTAA CTGCCCAATG AGTACTTCTT 420
GCCCCTGCC AAGATAGAGC CAGTTTACCA AGACAGGGGA ATTGCAGTAG AGAAAGAGTT 480
35 GAATATACAT AGAGCCAGCT AAATGGGAGA GTGGAGTTTT CTTATTACTT AAATCAGCCT 540
CCCYTAAAT TCAGAGGTGA GAATTTTCA AGGACAGTTT GGTGGSCAGG CCTAGGGAAT 600
40 GGATGCTGCT GATTGGCTAG GGATGCAATC ATAGGGGTGT AGAAAAGTWC CTTGTGCACT 660
GAGTCCACTT TTGGTGAGAG CTACCAAGGA GCTGCTGGTC TGCTGGTCCC GGTAGAGCCA 720
45 TCTGCTGTCA GGAATGCAAA AGTG 744

50 (2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 526 base pairs
(B) TYPE: nucleic acid
55 (C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

60 GCAGGGGAAT TCGCCACCG AGGGGTTTCA ACAGGGCTCG TGGGGTGAGG TGCAACACA 60

346

AAGCCCATAA GTGCTGGCCT GTTGGGACAA ATGAGAGAAA TCCCATAGGG TGGTGATGAC 120
 AGGCAATCA GCCATCTAT TCCCTGGGAA AATGAACCTT GTGCTCCTAT CAAATGCTCA 130
 5 GTTGTAAC TGGAAAAA TTTAGAAGA CTTCTGTCC AGCATCTGTG TTTATGTCTA 240
 TAAATGTAG AAACTAAG CACAGAGATG TTAATGTTT TGTCCAAGGT CCAACAGCTG 300
 GTTAGCAGC TTGCTCTGT CACCTTTCTA CTGAACCACA GTCCCGCTGG GCGAAGTCCT 360
 10 CAGCAGAT GGCTCTGCT ATAGCTGGG TATGGGAGT ATTAGTAGTT AACCAGTCAA 420
 CCAAGTTC CATAGCTAG GTTCTGCTC AGCTGGAGGT TAGGGAAAA CACAAGAAA 480
 15 TCCCTTACCA CTCTACAGT GCTGGGGAT GTACTAGAG ATCCCC 526

20 (2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

30 GGCACAGGGC AGGAGAGACT TGGTCCATGG GGAGAAGCCT GCAGTATAGA TGGGACCTCC 60
 AGGAGCCCAA GTAGCATAGA CCTGTCTGAT CCGGGGCCAT TGAGCCAGAG GATTTGGGCT 120
 35 GAATGTCCCC AGAGACAAA GGGAAAGGTA GATCCTTTCC CTAAAGATG AAAGCCATCG 180
 CCGGGGCTTG CTTATTGCTC TCTCTCCTGG TCCTTCCACA TGTGTCTCT GAACATTTGT 240
 TCTGGCATCA CAATCCCCGT CATCCTGTCA TCTGGCCCTT CCCACCTTTC CACCTTATCT 300
 40 CTTCAGTGT CTCGGGCTCG ACCTGGCACC TGGGTGAARG CTGCTCTTG CTGGTGCCCA 360
 TAGCCCCCAG TGTATGGTCT TGAMCTCCCC AGCCATATGG APACCCACCT CAGGAGGGCC 420
 45 CCTCGA 426

50 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 844 base pairs
 (B) TYPE: nucleic acid
 55 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

60 GGCACAGCGG CACGAGATAG GAAGCTTGGC AGGGCCAGCT CCCCAGTGC GCATTGCCCT 80

347

5 GTAACTCGAC CCGCTGGGAG TGGGGAGAGG CTTGGAAATG GAGCAGGGTG GTGGACCTCG 120
 TCTTCTCCTG CTGATCCCGAG GCCTCCTCCA TAACACCTAC CTAGCAGCGC CTGGGGACTT 180
 CCCAGCCCCA GGAACAACCTG AGAATACTGA GTGCCAGGGT AGCCCTAGCC CCATTTCACA 240
 CCTGGGCAAA GTGAGGTGAC TGGATTCAAA CACTCAGATT TAAACCTCCT CTGTGTCTGC 300
 10 AGCACCTGTA TATAACTGCC AGCCTCTGCT GCCCCTCTCC AAAAAGTCTC TGGCCCTTGT 360
 TTTGGCACCT GTCTCTGTCC TCCCCATTCT CTGCTCCTCC TTTCTCCAC TCAGANTCAC 420
 CCTGTTAGTT CAGCAAATGT TCATCGAGCT CCATAATGTA GCAGGACAGG NCTGTCTAAC 480
 15 AGATTCTGCG CTGCAAGGG TGAGACAAGT ACTCTCCATC TTTCTCTCAT CTTACAGAT 540
 GGTCTGCTCA ACAACTTTGC ACTGAATTGT AAATAATTGA TACTGCATAA AACATTGATG 600
 20 TTCTTTAAGG GTAGTCCAGC AAGGTGGCAA GTCTTATAAT GATAACTGCT CAAGGATCTC 660
 TCAGTGAAGC ATTTGGGGST GCTAGCTCTG CCTATGGGTG AGGTCAGCTA TCTCAGGCA 720
 TCTACTTCCA CTTGCCCGCC CATGCCAGGC TCACCTGAG CTGAGATGCC TGAGCAGGTG 780
 25 GCAGAAAGGA GCCACCTGGT TTATGCTTCG GGACCACAAA CTCCTCTATC CAGANGACAG 840
 TTTT 844

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1985 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

35 AGCCCTGCTG AAGTACAGGT TCTTCTATCA GTTCTGTTG GGCATGAAC GAGCAACAGC 60
 AAAGGAGATC AGGGATGAAT ATGTGGAGAC GCTGAGCAAG ATTTACCTGT CTTACTACCG 120
 45 CTCTTACCTG GGGCGGCTCA TGAAGGTGCA GTATGAGGAA GTGGCTGAGA AAGATGATCT 180
 AATGGGTGTG GAAGATACAG CAAAGAAAGG ATTCTYCTCA AAGCCATGCC TCCGCAGCAG 240
 50 GAACACCATT TTCACCCTAG GAACCCGCGG CTCTGTCATC TCCCCACTG AACTTGAGCC 300
 CCCCATCTG GTGCCTCACA CAGCCGAGCG GNAGAGCAGA GGTATCCATT TGAGGCCCTC 360
 TTCCGCAGCC AGCACTACGS CCTCCTAGAC AATTCTGCC GCGAATACCT TTTGATCTGT 420
 55 GAATTTTTTG TTGTGTCTGG CCCAGYTCCA CAGGACCTGT TCCATGCTGT CATGGGCCGT 480
 ACACTCAGCA TGACCCGTGA ACACCTGGAT TCTTATCTAG CTGACTGCTA CGATGCCATT 540
 60 GCTGTTTTTC TCTGTATCCA CATGTTTCTC CGGTTCCSTA ACATTGCAGC AAAGAGGGAT 600

348

GTTCCTGCCC TGGACAGGTA CTGGGGAACA GGTGCTTGCC TTGCTATGCC CACGCTTTGA 560
 ACTGATCCTG GAGATGAATG TTCAGAGCGT CCGAAGCACT GACCCCCAGC GCCTAGGGGG 720
 GTTGGATACT CCGCCCCACT ATATCACAGC CCGCTATGCA GAGTTCTCCT CCGCTCTTGT 780
 CAGTATCAAC CAGACAATTC CTAATGAACG GACCATGCAA TTGCTGGGAC AGCTGCAGGT 840
 GGAGGTGGAG AATTTTGTCC TCCGAGTGGC AGCTGAGTTC TCCTCAAGGA AGGAGCAGCT 900
 TGTGTTTCTG ATCAACAACT ATGACATGAT GCTGGGTGTG CTGATGGAGC GGGCTGCAGA 960
 TGACAGCAAA GAGGTTGAGA GCTTCCAGCA GCTGCTCAAT GCTCGGACAC AGGAATTCAT 1020
 TGAAGAGTTG CTGTCTCCCC CTTTGGGGG TTTAGTGGCA TTTGTGAAGG AGGCTGAGGC 1080
 TTTGATTGAG COTGGACAGG CTGAGCGACT TCGAGGGGAA GAAGCCCCGG TAACTCAGCT 1140
 GATCCGTGGC TTTGGTAGTT CCGGAAATC ATCAGTGGAA TCTCTGAGTC AGGATGTAAT 1200
 CCGGAGTTTC ACCAACTTCA GAATGGGAC CAGTATCATT CAGGGAGCGC TGACCCAGCT 1260
 GATCCAGCTC TATCATCGCT TCCACCGGGT GCTGTCCCAG CCGCAGCTCC GAGCCCTCCC 1320
 TGCCCCGGCT GAGCTCATCA ACATTACCA CTTTATGGTG GAGCTCAAGA AGCATAAGCC 1380
 CAACTTCTGA TGTGCCAGAA ACCGCCCTGA GATGTCCCGG TCATCTCCAT GGACTTCTGC 1440
 ACCCCATTCC ATACCCCTCT TCACCTGGGG TACCCCTTCC AGTTTCCGC TTGCTTCCCA 1500
 GGCCCTTGAC ATGGCTTACC TGCCCTCACT CCGAGCACCT TGCCCAACAG GATAAGCTGG 1560
 ATCCCTTTGG CTTTCTGAAT ATCCAGTGT CTTGAGGTT CCCAAGACCA CTTCCCTGTG 1620
 GGCTTCCAAA ATGGCCTTTA TCATTCTCC AGTCTGTAC CCTCCTTTCC TGCTCCCAT 1680
 CACCCAAGGC TTGTTTCTTC CCGTGTAAAA ACCACTGCCT CAATCTCTGG TTCACTCAAC 1740
 TAGTCACCAT GTCCTGAGGC ATGAAGCCTC CTCAGCTCTT GGAATTGCTG GCAAGGGGTG 1800
 ACTGCCTCTG AGTCATTGTG TTTTCAAG TGATTCTTT TCTGTAGCTT TTTGACCTAA 1860
 GATCTCAGCA APTTGAACAC TAACCTCTCC CCTCCTGGCT CAGAATTAC TCCGAAGTCA 1920
 GTCTGCAGAA AATAAATATT TAGTATGACA TGAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1980
 AAAAA 1985

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1416 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

ATATGAAGGG AAAGAATTGG ATTATGTTTT CTCAATTGAT GTCAATGAAG GTGGACCATC 60
5 ATATAAATTG CCGTATAATA CCACTGATGA CCCTTGGTTA ACTGCATACA ACTTCTTACA 120
GAAGAATGAT TTGAATCCTA TGTTCCTGGA TCAAGTAGCT AAATTTATTA TTGATAACAC 180
AAAAGGTCAA ATGTTGGGAC TTGGGAATCC CAGCTTTTCA GATCCATTTA CAGGTGGTGG 240
10 TCGGTATGTT CCGGGCTCTT CCGGATCTTC TAACACACTA CCCACAGCAG ATCCTTTTAC 300
AGGTGCTGGT CGTTATGTAC CAGGTTCTGC AAGTATGGGA ACTACCATGG CCGGAGTTGA 360
15 TCCATTTTACA GCGAATAGTG CCTACCGATC AGCTGCATCT AAAACAATGA ATATTTATTT 420
CCCTAAAAAA GAGGCTGTCA CATTTGACCA AGCAAAACCT ACACAATAT TAGGTAAACT 480
GAAGGAACCT AATGGAAGTG CACCTGAAGA GAAGAAGTTA ACTGAGGATG ACTTCATACT 540
20 TCTTGAGAAG AACTGTCTC TAATATGTAA TAGTTCTTCA GAAAAACCA CAGTCCAGCA 600
ACTTCAGATT TTGTGGAAG CTATTAACGT TCCTGAAGAT ATTGTCTTTC CTGCACTTGA 660
25 CATTCCTCGG TTGTCAATTA AACACCCAG TGTGAATGAG AACTTCTGCA ATGAAAAGGA 720
AGGGGCTCAG TTCAGCAGTC ATCTTATCAA TCTTCTGAAC CCTAAAGGAA AGCCAGCAAA 780
CCAGCTGCTT GCTCTCAGGA CTTTTTCGCA TTGTTTGTG GCCCAGGCAG GACAAAAACT 840
30 CATGATGTC CAGAGGGAAT CACTGATGTC CCATGCAATA GAACTGAAAT CAGGGAGCAA 900
TAAGAACATT CACATTGCTC TGGCTACATT GGCCTGAAC TATTCTGTTT GTTTTCATAA 960
35 AGACCATAAC ATTGAAGGGA AAGCCCAATG TTTGTCACTA ATTAGCACAA TCTTGGAAGT 1020
AGTACAAGAC CTAGAAGCCA CTTTATAGCT TCTTGTGGCT CTTGGAACAC TTATCAGTGA 1080
TGATTCAAAT GCTGTACAAT TAGCCAAGTC TTTAGGTGTT GATTCTCAA TAAAAAGTA 1140
40 TTCTCAGTA TCAGAACCAG CTAAAGTAAG TCAATGCTGT AGATTTATCC TAAATTTGCT 1200
GTAGCAGTGG GGAAGAGGGA CGGATATTTT TAATTGATTA GTGTTTTTTT CCTCACATTT 1260
45 GACATGACTG ATAACAGATA ATTAAAAAAA GAGAATACGG TCGATTAAAT AAAATTTTAC 1320
ATCTTGTAAG GTGGTGGGGA GGGGAAACAG AAATAAAATT TTTGCACTGC TGAAAAAAA 1380
50 AAAAAAAA AAAAGGAAAC TCGAGGGGGG GCGCGG 1416

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1935 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

5	NTCTACCCCTA ATCAAGATGG GGACATACTT CGCGACCAGG TTCTTCATGA ACATATCCAG	60
	AGATTGTCTA AAGTAGTGAC TGCAAATCAC AGAGCTCTTC AGATACCAGA GGTTCATCTT	120
	CGAGAAGCAC CATGGCCATC TGCACAATCA GAAATCAGGA CAATAAGTGC TTATAAAACC	180
10	CCCCGGGACA AAGTGCAGTG CATCCTGAGA ATGTGCTCTA CGATTATGAA CCTCCTGAGC	240
	CTGGCCCAATG AGGACTCTGT CCGTGGAGCG GATGACTTTG TTCCTGTGTT GGTGTTTGTG	300
	TTGATAAAGG CAAATCCACC CTGTTTGCTG TCTACTGTGC AGTATATCAG TAGCTTTTAT	360
15	GCTAGCTGTC TGTCTGGAGA GGAGTCCTAT TGGTGGATGC AGTTCACAGC AGCAGTAGAA	420
	TTCAATTAATA CCATCGATGA CCGAAAGTGA CCAAGACCAA GGGCCACCAA GGCAGCAGAC	480
20	TGTTAATCAG ACAAACAGAT CTCTGAGAAG GTGCATCAGC TGCTTTGAAG GCTGAAGATT	540
	GTTTTGTATG ATRACTGCACA GCATCAGGCA TTTTAAAGCA GATCTTTACT AAACAGGTTA	600
	ATGAGCTAAC AAGCAGGTTT TCTCGTCTTT GGGCTCTTTT CTTTCTGAGT TGCATATTCT	660
25	ATTTTCTTGT CCCCAAGTAG AGACTAGTAG TACAAAAAGG GACCACATTT TTCAAGTATT	720
	TCTAAGTATA AAAAACAAAA CAAAAATCTC TTAGGAAATG TCTAGACCTC CATTCTTGGA	780
30	TTCCCTTTCT TTCCTTTTAT TTAAAAAAG AACAGTACCC CTCTTTTAAG ATGCTGTCTT	840
	ACATTAAATGA GCATCTAATG GAAAGAAGGT ATGAGTTGCA CTGAGGATTA GAATAGTGGT	900
35	GCGTTAGTGG CATTATCTAT AAATACACTC ACCTAAATTG AAAGCTAAGA AGGAAATGTA	960
	AATATAATAT ATATTTATAT TTGATGTAAT ATGGACATCT GCAGATTCTA ATAAACAAGG	1020
	ACTATTGCTG ATAGTAGGCT GTGACATACT GTCTGTGAA ATGGTTTCCT TGACAAAATT	1080
40	TAAGCTGAGC TTAAGACCA AAAAACAAAA AGTACACAGA AATATTTATT AAAATGTAAT	1140
	ACAGTTTATT GAACTTTCTA GGTATGGAGT TTGATGGACA GGGCTGCCTY TAATGAGTGT	1200
45	GAAGGTCACT AAGTCACTTA GACATCTCAC CGTGGAAAGT TGTGAGCCTG CATTAGGAGA	1260
	TAGACTGATT ACCATACATG ACATAAAAAG GAACAGTGGG TAGCTCATAC TTTATGGTGG	1320
	TTCTTCTCCT CCGAAATAAT ATACTGCAGA AATCCAGAC AGAGCTCCTT ACAAAACCTT	1380
50	AATTGTAATA TATTTTGTAT GATTATTCAC ATTGAATGCA CAGACCAAGA ATTCAGTGAA	1440
	TGTCATTTTT TAAAAAATA ATTTGTATTG TCTGCTCTAG TGATACAAGT TTTACTAGTG	1500
55	ATAAACTATT TTAATCAACC ATACTATTCT TATGGAAAAA AATATCTATT TTGCCAGGTT	1560
	TCTGTGCCTT TATTTCCCTC TTCTGAAAAA AAGTCTGTGT TTTCATAGTT TGGTTTGCAT	1620
	TGTATATCAA TAATTAATCA GGAATGGGTT TTGGTGCCTG AAAAATGGC CATGGAGGCA	1680
60	CACCAAGCT TCAAGCACAA GTCTTGTACA TGGGCCATCA CTGTCTGGTT TCACTTCGTG	1740

TGTTCCTAA ACACATTTAG CTGCTTTTTT AACAACTCA GCCCCATCT TGAGTCCCTT 1800
 GTTGTGGGA GCATTTCCAG GCATTTTTTA AGGGAAGTGT GACAAACAGC CTCGGGCAGA 1860
 TGAACACCGA GGCTCTCTGT TGTCTGTCTC TGAGATCTTT GTGTCTGGGA ATGCCTAAAG 1920
 NTTTTGNTTT TTTTT 1935

10

(2) INFORMATION FOR SEQ ID NO: 100:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 599 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GAATTCGGCA CGAGCGTCCA CGCAGCCGCC GGGGGGCCAG CACCCAGGCC CCTGCATGCC 60
 AGGTCGTTGG AGGTGGCAGC GAGACATGCA CCGGGCCCGG AAGCTCCTCA GCCTCCTCTT 120
 CCTCATCCTG ATGGGCACTG AACTCACTCA AGACTCCGCT GCCCCGACT CCCTGCTGAG 180
 AAGTTCAAAG GGCAGCACGA GGGGGTCTTT GGCTGCTATT GTCATCTGGA GGGGAAGAG 240
 TGAGAGCCCG ATAGCCAAGA CCCCAGGCAT TPTCAGAGGT GCGGGGACCT TAGTCCTACC 300
 CCCAAGACAC ACCCCTGAGT GGCTCATCCT CCCTTTGGGC ATAACGCTGC CCTTGGGGGC 360
 TCCAGAAACA GCGGTGGGG ATTGTCCCGC TGAGACCTGG AAGGCAGCC AGCGTGCCCG 420
 CCAGCTGTGT GCATGCTGG CTTAATATGC AGGCCTTGGG GGGCTGTGGC CACATGCCCG 480
 GCAGGAGGTG AGTGAGGAGC CCTGTGCCGT GGTGGTGTGG GATCGTGGG CATTTCAAAC 540
 GGGCTTGTGG TACCCTGAAC AATGTATCAA TAGAGAAAAA AAAAAAAAAA AAAACTCGA 599

45

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 734 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

55

GAATTCGGCA CAGAAAAAAA AGAGAGACTG GGTCTTACTG TGTGCCCAG ACTTGTCTTG 60
 AACTCCTGCC TCAGCCTCTC AAGTACTTGG GATTATAGGC CAAGAAGCCA CCATGCCTAG 120
 CTTCTTCTGT TCATTGATCC AGACTAATAC TCTGGGTCA GCCTCATTTT TTCTCTTTCT 180

60

	CACTTTGCAC ATCCACTTGT CACCAAATCK RGTTCATTCT GCATCCTAAG TAAGTCCTTT	240
5	GATTCCCTCCA GTPGTTCAAT AGTAATGTCT CAARTGTAAT TTTTCTAGT AGTTTTCAGC	300
	CTGTCTTTCC KGCCTTCAGT CTTAACTTCT CCAGTACATA KCCACATTG TTGTCAGCAK	360
	GATCAWATTT TATTTAAAAA TACTTTACAW AKGTTTATKG CCAAATATTA GRAAATACAG	420
10	ATTCATGGAA AGAAAAATCA CTGTCCCAAG GAGGTCACCTG GCATCGTGAG GTTAAGGGGT	480
	GATTTTAATT TTAAAAATG TATATTTTTT CCTGTGTAGA GTAGTAACAC CCTTGAAAAC	540
15	ACAWTCCCTT GTAAAGTCTC TAATTCTGTA CTCCGCATCT AGSTGRTCTC TTCTTTCTCA	600
	GATATTTTAC AATTTCAATT ATCACCACCT TTCTCTAGCC TTTACCCGTC TCTTCAATAT	660
	TWACATATGC AGAAGTTTCT CCTAACAAAC ACCTGCCTCT GCCTCAGTTC TGCTACCACC	720
20	CTGTGCTTT CTTCCTCTC ACAATCAAAT TTAAGAGTGT CAAAAAATA AAAAAAATC	780
	TCGA	784

25

(2) INFORMATION FOR SEQ ID NO: 102:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1035 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

40

	AGAGGCCTGG CTGCGTTGCC CTATCTCCGT CTCCGCCACC CACTTAGCGT TTTAGGCATC	60
	AATTACCAGC AGTTTCTCCG CCACTATCTG GAAAAATTACC CGATTGCTCC CGGCAGAATA	120
	CAAGAGCTTG AAGAAGCCCG CAGTTGCGTG GAAGCCTGCA GAGCAAGGGA AGCAGCGTTT	180
	GATGCCGAAT ATCAGCGAAA TCCTCACAGG GTGGACCTCG ATATTTTAAC CTTTACGATA	240
45	GCTCTGACTG CCTCTGAAGT TATCAACCCT CTGATAGAAG AACTTGCTTG CGATAAGTTT	300
	ATCAATAGAG AATAGTTAGG TGGTGACACT ACTTCAAGAG AACCTCTGCA TTCCAGTCAT	360
50	ACCAATCCTG CAACTTGATT TTCAGAAGTC AAGAGTATAT CGCGATAAGA CAGTGCAACG	420
	GTGGAGGGGA AAAAAAGGGG GAGGGGGAAG CTTATCTTGA AAAAGCATCA CAGAAGTAGA	480
	AAAAATGTC GAAAGCAATTA TAACTGTAAC GTTCTTTGAG TTTGTGATTG ATCCACATTT	540
55	TTCCCCCTGC ATTATGGAAA ATGTCTCTCA GCATTGCTTT ATTACAAAGT AAAGGATGGT	600
	TTTATAAAAT TGAGACTGAT GAAACATCAA TACTAGAGCC CATGAGGATG AAAGAAATTA	660
60	TCAATAGTG CTGAACAGAA TAAGATGTTA ACGCTGAGTT ATTAGGACTG GAAGGCTATG	720

353

AAAAGAACTT GAAATTGTGG GAATATGTGC TCTCTTCATG TCATATTCAA TAGAAGTTTC 780
 TAGTTTAAGA TTGATTTTGT GTTTTCTTAG GCATTTCAAG TCACAAGCAA AGTAAATGTA 840
 5 TATATTATGT GATAAATCAT GTTTTCAAGA ACGTCAAATT TCTGGACTTT TTTCTTTCAA 900
 TTTTAAATTT TTAAGTTTTT TTTGGTATTA AAAAATCTAT TCACAAGCCA AAAAATWTWT 960
 WAAATWTWCM GCGAAAAGCC AAAAAAAAAA AAAAMMAGGG GGGGCGGGG CCGATCCCCC 1020
 10 CAAGGGGGTC CNGMT 1035

15

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2218 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

AGGTATTAGG CCGTTTTGTG GGAGCCCAT GTTTTGTGTTT TCTGAGTTGG TGGGGAGGGA 60
 SGGAGGGGGA GGGCTGAATT GTTTTCAGA GGAAGATGCC ATCTGTGCTT TAAATTTCTC 120
 30 ATTACTGGGT TAGAAAACAA AGAGGGAKTG CCGTGCACAT TTTCTTTTGT GCTTTTAAAT 180
 GTTCTTAAG TTGGAACAGG TTTCCTCGGG CCGTGTGGA CTGATTGCTG GAGTGCATTT 240
 GATAGTTAAA AATTACTAAT TGGTTTTATT TCCCTTCACA CTCTGCCTCC CCACTTCTCC 300
 35 CCGGTTACT GAAAAATAAC CATTTTAGTG TCAGGCTAGA AATTGAATTG CTGAGTTTGT 360
 TGTATCCTTT AAATTAAAA CCACAAGTGT TTATTGTAGT GGTAAACTG TAGCATCTCA 420
 40 GCATCTGGGT GGAAGCTGCC TATATTCTT CCCAGTTTAA CTGGGGACCA TCTGTGAAAT 480
 TAATTTTCCA TCCAGACAGC TGCTGTGAGC AAATGAACAT AAATGCTGCC TGGAAATTTA 540
 CTAACCAATT TTTATATTGA CCGCAGTGT AAAAAGCACA TTAAATTATA AACAAATAT 600
 45 TCAAAATGGG CAAATTTTAT TTTCAAATGC AGTGTAGAGC TAGATTAAAA GCAACTCTTT 660
 GCCACCTACT CTGCCCTTTT GGCAAAGTTA CTTTGAACAA AGAATCTTAA GGGTTTATTA 720
 50 AGAACTCTTT ATTTTCTTCA TACCCTGTTT TCTGCAGTGC TTTCTAACAG CTTCTGGGTG 780
 CAGATTTTCT TCGGCATCCT TTGCACTCA GCTTATTACA GGTAGGTAGT GCTTAAGAAA 840
 AGTCATGGAG GACTAAAGCC TAAGTCTTTT TCACTTTTCC TCCATCTGAA GGTAGGTGAG 900
 55 TTCACTCTCT TCATAGTAAT GCTGTTTAC CAAGACTTTA TAGCAGATGG ACCCAGAAAG 960
 AATTTTCTGC TATTGTGTTT ACTACAACAG GATAGGGACA TCAGACAGCC CCAGAAACCC 1020
 60 CTTCCAGATC TGATATGGGA CTATTAAATT TTATGCTGTT AATTGGTATT CATTCACAAT 1080

GCAGTTGAAG GGGGAAGGCT CCACTGCATT CTTTGGCTAA GGCCTGAATG CTIGCTCATC 1140
 5 TGTAAGATCT ATACTCGAGG TTTTGTTC TCMTTAAAT TCTTTAGGGA GAGAGGGATG 1200
 GTTCTCGAGG GGTCTGAAA GTATGATTCA ATGTGCAACA TACAGGTAGG TCTTCAGCAT 1260
 AAGCTGAAAT ATATGCATGT AAAAAGTTTG ACATCTTTTT TTTTAATTTT CCACTTTCTT 1320
 10 CTTAACTTTA CTCTCTTTTT TGTCCCCCCC CCATCTTACA GAAGTTGAGG CCAAGGGAGA 1380
 ATGGTAGGCA CAGAAGAAAC ATGGCAAACT GCTCTGTGCT TTCAAACCAA AGTGTTCCTC 1440
 15 CCAACCCCAA ATTTGTCTAA GCACTGGCCA GTCTGTTGTG GGCATTGTTT TCTACAACCA 1500
 AATTCTGGGT TTTTTCTTC TTTCTTAAA CATAGAGGTA CCACCACAAG GCATGCCCTA 1560
 CTCTCTGCA GCTCTGAAA GCATCTGTTT GAGGGAAGG TCTCTGGCA ACCAAGTGGT 1620
 20 TATTGGATT GCTGCTTCC CTTTTCCAC CTGGGACATT GYAATCATAA AATAACAGTA 1680
 AATTCCAAAC CTCAAAACT ATTATGGCT GAGCAGCT GAAATCTAGC AGAGTTTAC 1740
 TCTCTGCCT CCATGTCTGT CACTTATAAT TCAGGTTCTG CTGTTGGCTT CAGAACATGA 1800
 25 GCAGAAGAAT CTTTTATGC TAGTTATTGC ATTCATGGTT GAAACTCAAC TTAGGGAAAG 1860
 GGTCCAATG TATTAAGCAA TGGGCTGCTT CTCCTCAATC CTCCTAACA ATTCGTTCTG 1920
 30 TGGACTTCTC ATCTAAAAGG TTAGTGGCTT TTGCTTGGGA TCAGTGTCTT CTATTGATGT 1980
 TCTGCTGGT CTCCAGACAC ATTCCTGTTG CATTAGACT TGAAAGACTT GTAGATGTGT 2040
 GATGTTGAG CACAGGATGC TGAAAGCTAT GTTACTATTC TTAGTTGTA AATTGTCCTT 2100
 35 TTGATACCAT CATCTGTTT TCTTTTGTG GGTATAAATA AAAACACTGT TGACAATAAA 2160
 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2218
 40

(2) INFORMATION FOR SEQ ID NO: 104:

- 45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1351 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

CTTACAGAC TGACAGAATG GTTTGTTTT GTTTGTTTT GTTTGTTTT GTTTTGAGA 60
 55 TGGACTCTAG CTCTGTACC CAGGCTGGAG TGCAGTGGTG CGATCTGGC TCACTGCAAG 120
 CTCGGCTCC CGGTTCTCA CATTCTCTT GCCTCAGCT CCGAGTAGC TGGGACTACA 180
 60 GCGGCCCCC ACCACGCCC GCTAATTTTT TGTATTTTT AGTAGAGACG GGGTTTCACC 240

	ATGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCGTGATC GCGCCGCTTC GGCTCCCAA	300
	AGTGCTGGGA TTACAGCGGT GAGCCACCGT GCGTCCCCCA GAATGGTTTT TAAAGCCACA	360
5	GTTGAGARGC CACCCATTGC CCGCGCGCTG GACAGTGATC ATCTTGTTCA TCTTGTTGAG	420
	TCCTTTCTTG TGTGATTGGA ATTATTCATC CCCTTTGAAA GATGAGAAGG TTGAGATGCA	480
10	AAGACTCTAC CTTTCCAAGT TCTCACTGCT GGAAAGARCT AGAAGCACAG TTCAAAAGTTC	540
	TGNNPTCTGG ACTCTGCAGT CCAGGTYTCC CTNPTCCAC TTGCCTACCC TCAATGCCAC	600
	ACTGTTTTTG AAGTGGCCCA TAAGTTGAAG GRAAAGTTTA AAGACAGTTC AATTTAATCA	660
15	TCAGRATGCA TTCTTTTTT TTTCGGARAC GGAATTTTAC TCTTGCTGCC CAGCTGGAG	720
	TGCAATGCTG CAATGATCTC GGCTCACTGC AACCTATGCC TCCTGGGTTT AAGNGATTAT	780
20	CCAGCCTCAG CCTCCCGAGT AGCTGGGATT ATGGGCGCCC ACCACCATGC CCAGCTAAT	840
	TTTGTATTTT TTTTTTTAGT AGAGATGGGG TTTCGCCAGG TTGGCCAGGC TGKTCTTGTG	900
	AATCTCTGGC YTCAGGTGAT YTGCCACAT CATCTTCCAA AAGTGCTGGG ATTACAGGCA	960
25	TGAGCCACTG CGCCTGGCYT CAGAATGCAT TCTTACACAT CTATCCTAGA CATTTATAAG	1020
	CACTCTAATG GATAACAATC CAAGAATAAA TGATTGTAAA AGATGATGCC GAACAGTTGA	1080
30	TGTCAATCTT TTTTTCCTAA GAAAAAAGT CCGCGAGTAT TAAATATTTA GATCAATGTT	1140
	TATAAANTGA TTAATTTGTA TATCTCATTA TTCCTATTTT GGAATAAAAA CTGACCTTCT	1200
	TTAATCATAT ACTTGCTTTT TGTAATAGC AGCTTTTGTG TCATTCTCCC CACTTTATTA	1260
35	GTTAATTTAA ATTGGAAAA ACCCTCAAAC TAATATCTT GTCTGTTCCA GTCTTATAAA	1320
	TAAACTTAT AATGCATGTA AAAAAAAAAA A	1351

40

(2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2066 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

	GGCAGGAGGC GCGGAGGGG CACAATCACA GCTCCGGGCA TTGGGGGAAC CCGAGCCGGC	60
	TGCGCGGGG GAATCCGTGC GGGCGCCTT CGTCCCGGTC CCATCCTGCG CCGGCTCCAG	120
55	CACCTCTGAA GTTTTGCAGC GCGCAGAAAG GAGGCGAGGA AGGAGGGAGT GTGTGAGAGG	180
	AGGGAGCAAA AAGCTCACCC TAAACATTT ATTTCAAGGA GAAAAGAAAA AGGGGGGGCG	240
60	CAAAAATGGC TGGGGCAATT ATAGAAAACA TGAGCACCAA GAAGCTGTGC ATTGTTGGTG	300

	GGATTCTGCT CGTGTTCCAA ATCATCGCCT TTCTGGTGGG AGGCTTGATT GCTCCAGGGC	360
5	CCACAACGGC AGTGTCTTAC ATGTCCGTGA AATGTGTGGA TGCCCGTAAG AACCATCACA	420
	AGACAAAATG GTTCGTGCCT TGGGGACCCA ATCATTGTGA CAAGATCCGA GACATTGAAG	480
	AGGCAATTCC AACGGAAATT GAAGCCAATG ACATCGTGT TTCTGTTTAC ATTCCCTCC	540
10	CCCACATGGA GATGAGTCCT TGGTTCCAAT TCATGCTGTT TATCCTGCAG CTGGACATTG	600
	CCTTCAGCT AAACAACCAA ATCAGAGAAA ATGCAGAACT CTCCATGGAC GTTTCCTGG	660
15	CTTACCGTGA TGACGCATTT GCTGAGTGA CTGAAATGGC CCATGAAAGA GTACCACGGA	720
	AACTCAATG CACCTTCACA TCTCCCAAGA CTCCAGAGCA TGAGGGCCGT TACTATGAAT	780
	GTGATGTCCT TCCTTTTCATG GAAATGGGT CTGTGGCCCA TAAGTTTAC CTTTTAAACA	840
20	TCCGGCTGCC TGTGAATGAG AAGAAGAAAA TCAATGTGGG AATTGGGGAG ATAAAGGATA	900
	TCCGGTTGGT GGGGATCCAC CAAATGGAG GCTTCACCAA GGTGTGGTTT GCCATGAAGA	960
25	CCTTCCTTAC GCCCAGCATC TTCATCATTA TGGTGTGTA TTGGAGGAGG ATCACCATGA	1020
	TGTCCCGACC CCCAGTGCTT CTGGAAAAAG TCATCTTTGC CCTTGGGATT TCCATGACCT	1080
	TTATCAATAT CCCAGTGGAA TGGTTTCCA TCGGGTTTGA CTGGACCTGG ATGCTGCTGT	1140
30	TTGGTGACAT CCGACAGGGC ATCTTCTATG CGATGCTTCT GTCCCTCTGG ATCATCTTCT	1200
	GTGGCGAGCA CATGATGGAT CAGCAGGAGC GGAACCACAT TGCAGGGTAT TGAAGCAAG	1260
35	TGGGACCCAT TGCCGTTGGC TCCTTCTGCC TCTTCATATT TGACATGTGT GAGAGAGGG	1320
	TACAACTCAC GAATCCCTTC TACAGTATCT GGAATACAGA CATTGGAACA GAGCTGGCCA	1380
	TGGCCTTCAT CATCGTGGCT GGAATCTGCC TCTGCCTCTA CTTCCTGTTT CTATGCTTCA	1440
40	TGGTATTICA GGTGTTTGG AACATCAGTG GGAAGCAGTC CAGCCTGCCA GCTATGAGCA	1500
	AAGTCCGGCG GCTACACTAT GAGGGGCTAA TTTTATAGTT CAAGTTCCTC ATGCTTATCA	1560
45	CCTTGGCCTG CGCTGCCATG ACTGTCATCT TCTTCATCGT TAGTCAGGTA ACGGAAGGCC	1620
	ATTGGAAATG GGGCGGCGTC ACAGTCCAAG TGAACAGTGC CTTTTTCACA GGCATCTATG	1680
	GGATGTGGAA TCTGTATGTC TTTGCTCTGA TGTCTTGTG TGCACCATCC CATAAAAACT	1740
50	ATGGAGAAGA CCAGTCCAAT GGAATGCAAC TCCCATGTAA ATCGAGGGAA GATTGTGCTT	1800
	TGTTTGTTC GGAACTTTAT CAAGAATTGT TCAGCGCTTC GAAATATTCC TTCATCAATG	1860
55	ACAACGCAGC TTCTGGTATT TGAGTCAACA AGGCAACACA TGTATTACAG CTTTGCATTT	1920
	GCAGTTGTCA CAGTCACATT GATTGTACTT GTATACGCAC ACAAATACAC TCATTAGCC	1980
	TTTATCTCAA AATGTTAAAT ATAAGGAAAA AAGCCTCAAC AATAAATATT CTTGAGTATA	2040
60	AAAAAAAAA AAAAAAAAAA AAAAAA	2066

5 (2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1705 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15 AATCGGCAK AGGGCAGCTG TCGGCTGGAA GGAAGTGGTC TGCTCACACT TGCTGGCTTG 60
CGCATCAGGA CTGGCTTTAT CTCTGACTC ACGGTGCAAA GGTGCACTCT GCGAACGTTA 120
AGTCCGTCCC CAGCGCTTGG AATCCTACGG CCCCCACAGC GGGATCCGCT CAGGCTTCCA 180
20 GGTCTCAAC TCCTGYGGAC GGTGAACAAT GGCCTCCATG GGGCTACAGG TAATGGGCAT 240
CGCGCTGGCC GTCTGGGCT GGTGGCCGT CATGCTGTGC TCGCGCTGC CCATGTGGCG 300
25 CGTGACGGCC TTCATCGGCA GCAACATTGT CACCTCGCAG ACCATCTGGG AGGGCCTATG 360
GATGAACTGC GTGGTGCAAG GCACCGGCCA GATGCAGTGC AAGGTGTACG ACTCGCTGCT 420
30 GGCAGTGGCG CAGGACCTGC AGGCGGCCCG CGCCCTCGTC ATCATCAGCA TCATCGTGGC 480
TGCTCTGGGC GTGCTGCTGT CCGTGGTGGG GGGCAAGTGT ACCAACTGCC TGGAGGATGA 540
AAGCGCCAAG GCCAAGACCA TGATCGTGGC GGGCGTGGTG TTCCTGTTGG CCGGCCTTAT 600
35 GGTGATAGTG CCGGTGTCTT GGACGGCCCA CAACATCATC CAAGACTTCT ACAATCCGCT 660
GGTGGCCTCC GGGCAGAAGC GGGAGATGGG TGCCTCGCTC TACGTGGGCT GGGCCGCCTC 720
CGGNTGCTCT CTCCTTGGCG GGGGCTGCT TTGCTGCAAC TGTCCACCCC GCACAGACAA 780
40 GCCTTACTCC GCCAAGTATT CTGCTGCCCG CTCTGCTGCT GCCAGCAACT ACGTGTAAAG 840
TGCCACGGCT CCACTCTGTT CCTCTCTGCT TTGTTCTTCC CTGGACTGAG CTCAGCGCAG 900
45 GCTGTGACCC CAGGAGGGCC CTGCCACGGG CCACTGGCTG CTGGGGACTG GGGACTGGGC 960
AGAGACTGAG CCAGGCAGGA AGGCAGCAGC CTTACAGCTC TCTGGCCCAC TCGGACAACT 1020
TCCCAAGGCC GCCTCTGCTT AGCAAGAACA GAGTCCACCC TCCTCTGGAT ATTGGGGAGG 1080
50 GACCGAAGTG ACAGGGTGTG GTGGTGGAGT GGGGAGCTGG CTTCTGCTGG CCAGGATGGC 1140
TTAACCCCTGA CTTTGGGATC TGCTGCAATC GGTGTTGGCC ACTGTCCCCA TTTACATTTT 1200
55 CCCCAGCTGT TCTGCTGCA TCTCTCTGT TCGGGTAGG CCTGATATC ACCTCTGGGA 1260
CTGTGCCTTG CTCACCGAAA CCGCGGCCCA GGAGTATGGC TGAGGCCTTG CCCACCCACC 1320
60 TGCCTGGGAA GTGCAGAGTG GATGGACGGG TTTAGAGGGG AGGGGCGAAG GTGCTGTAAA 1380

358

CAGGTTTGGG CAGTGGTGGG GGAGGGGGGC AGAGAGGGCG CTCAGGTTGC CCAGCTCTGT 1440
 GGCCTCAGGA CTCTCTGCCT CACCCGCTTC AGCCCGAGGC CCTGGAGAC TGATCCCCCTC 1500
 5 TGAGTCCTCT GCCCCTTCCA AGGACACTAA TGAGCCTGGG AGGCTGGCAG GGAGGAGGGG 1560
 ACAGCTTCAC CCTTGAAGT CTTGGGGTTT TTCTCTTCC TTCTTTGTGG TTTCTGTTTT 1620
 10 GTAATTTAAG AAGAGCTATT CATCACTGTA ATTATTATTA TTTTCTACAA TAAATGGCAG 1680
 CTGTGCACAG GRAAAAAAAA AAAAG 1705

15

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1167 base pairs
 20 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TGCAGGAATT CGGCAGAGGT TTTCCGCTAG ACTCTGGCAG TTGGTGAGCA TCATGGCAAC 60
 CGTTACAGCC ACAACCAAAG TCCCGGAGAT CCGTGATGTA ACAAGGATTG AGCGAATCGG 120
 30 TGCCCACTCC CACATCCGGG GACTGGGGCT GGACGATGCC TTGGAGCCTC GGCAGGCTTC 180
 GCAAGGCATG GTGGGTACAG TGGCGGCACG GCGGGCGGCT GGCCTGGTGC TGGAGATGAT 240
 CCGGGAAGGG AAGATTGCCG GTCGGGCAGT CCTATTGCT GGCCAGCCCG GCACGGGGAA 300
 35 GACGGCCATC GCCATGGGCA TGGCGCAGGC CCTGGGCCCT GACACGCCAT TCACAGCCAT 360
 CGCCGGCAGT GAAATCTTCT CCCTGGAGAT GAGCAAGACC GAGGCGCTGA CGCAGGCCTT 420
 40 CCGGCGGTCC ATCGGCGTTC GCATCAAGGA GGAGACGGAG ATCATCGAAG GGGAGGTGGT 480
 GGAGATCCAG ATTGATCGAC CAGCAACAGG GACGGGCTCC AAGGTGGGCA AACTGACCC 540
 45 CAAGACCACA GAGATGGAGA CCATCTACGA CCTGGGCACC AAGATGATTG AKTCCCTGAC 600
 CAAGGACAAG GTCCAGGCCG GGGACGTGAT CACCATCGAC AAGGCGACCG GCAAGATCTC 660
 CAAGCTGGGC CGCTCCTTCA CACGCGCCCG CGAACTACGA CGCTATGGGC TCCCAGACCA 720
 50 AGTTCGTGCA GTGCCCAGAT GGGGAGCTCC AGAAACGCAA GGAGGTGGTG CACACCGTGT 780
 CCTTGCACGA GATCGACGTC ATCAACTCTC GCACCCAGGG CTCTCTGGCG CTCTTCTCAG 840
 GTGACACAGG GGAGATCAAG TCAGAACTCC GTGAGCAGAT CAATGCCAAG GTGGCTGAGT 900
 55 GCGCGGAGGA GGGCAAGGCG GAGATCATCC CTGGAGTGCT GTTCATCGAC GAGGTCCACA 960
 TGCTGGACAT CGAGAGCTTC TCCTTCTCA ACCGGGCCCT GGAGAGTGAC ATGGGCGCTG 1020
 60 TCCAGCAGGT CTATGGGGAT GCGGTGAGCG CTCTGGTAGC TGGTGCCCGG GATTCCGGTG 1080

ATGCCACGGT TGGTGGGCTC GTGCCGAATT CCTGCGCCG GGGGGATCC CTACTTCTAG 1140
 AGCGGGCGCC ACCGCGGTGG ANCTCCN 1167

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1907 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

GGCACAGGGG AATCATCGTG TGATGTGTGT GCTGCCTTTG TGAGTGTGTG GAGTCTGTCT 60
 CAGGTGTTAG GTACAGTGTG TTTGATCGTG GTGGCTTGAG GGAACCCCTT GTTCAGAGCT 120
 GTGACTGCGG CTGCACTCAG AGAAGCTGCC CTTGGCTGCT CGTAGCGCGG GGCCTTCTCT 180
 CCTCGTCATC ATCCAGAGCA GCCAGTGTCC GGGAGGCAGA AGGTACCGGG GCAGCTACTG 240
 GAGGACTGTG CCGGCCTGCC TGGGCTGCCC CCTCGGCCGT GGGGCGCTGT TGCTGCTGTC 300
 CATCTATTTT TACTACTCCC TCCCAAATGC GGTGCGCCCG CCCTTCACTT CGATGCTTGC 360
 CCTCCTGGGC CTCTCGCAGG CACTGAACAT CCTCCTGGGC CTCAGGGGCC TGGCCCCAGC 420
 TGAGATCTCT GCAGTGTGTG AAAAAGGGAA TTTCAACGTG GCCCATGGGC TGGCATGGTC 480
 ATATTACATC GGATATCTGC GGCTGATCCT GCCAGAGCTC CAGGCCCCGA TTCCAACTTA 540
 CAATCAGCAT TACAACAACC TGCTACGGGG TGCAGTGAGC CAGCGGCTGT ATATTCTCCT 600
 CCCATTGGAC TGTGGGGTGC CTGATAACCT GAGTATGGCT GACCCCAACA TTCCCTTCCT 660
 GGATAAACTG CCCCAGCAGA CCGGTGACCG TGCTGGCATC AAGGATCGGG TTTACAGCAA 720
 GAGCATCTAT GAGCTTCTGG AGAACGGGCA GCGGGCGGGC ACCTGTGTCC TGGAGTACGC 780
 CACCCCTTGG CAGACTTTGT TTGCCATGTC ACAATACAGT CAAGCTGGCT TTACGGGGGA 840
 GGATAGGCTT GAGCAGGCCA AACTCTTCTG CCGGACACTT GAGGACATCC TGGCAGATGC 900
 CCCTGAGTCT CAGAACAACCT GCGGCTCAT TGCCTACCAG GAACCTGCAG ATGACAGCAG 960
 CTTCTCGCTG TCCCAGGAGG TTCTCCGGCA COTGCGGCAG GAGGAAAAGG AAGAGCTTAC 1020
 TGTGGGCAGC TTGAAGACCT CAGCGGTGCC CAGTACCTCC ACCATGTCCC AAGAGCCTCA 1080
 GCTCCTCATC AGTGGAAATG AAAAGCCCCC CCCTCTCCGC ACCGATTTCT CTTGAGACCC 1140
 AGGCTCACCA GCCCAGAGCC TCCAGTGGTC TCCAAGCCTC TGGACTGGGG GCTCTTTCA 1200
 GTGGCTGAAT GTCCAGCAGA GCTATTTCTT TCCACAGGGG GCCTTGCAAG GAAGGGTCCA 1260

360

5 GGACTTGACA TCTTAAGATG COTCTTGTCG CCTTGGGCCA GTCATTTCCT CTCTCTGAGC 1320
 CTCGGTGTCT TCRACCTGTG AAATGGGATC ATAATCACTG COTTACCTCC CTCACGGTTG 1380
 10 TTGTGAGGAC TCACTGTGTG GAAGTTTTTC ATAAACTTTG GATGCTAGTG TACTTAGGGG 1440
 GTGTGCCAGG TGTCTTTTCAT GGGGCTTCC AGACCCACTC CCCACCTTC TCCCCTTCCT 1500
 TTGCCCCGGG ACGCCGAAT CTCTCAATGG TATCAACAGG CTCCTTCGCC CTCTGGCTCC 1560
 TGGTCATGTT CCATTATTGG GGAGCCCCAG CAGAAGAATG GAGAGGAGGA GGAGGCTGAG 1620
 TTTGGGGTAT TGAATCCCCC GGCTCCCACC CTGCAGCATC AAGGTTGCTA TGGACTCTCC 1680
 15 TGCCCGGCAA CTCCTGCGTA ATCATGAGTA TCTCTAGGAT TCTGGCACCA CTCCTTCCC 1740
 TGGCCCTTA AGCCTAGCTG TGTATGGGCA CCCCCACCCC ACTAGAGTAC TCCCTCTCAC 1800
 TTGCGGTTTC CTTATACTCC ACCCCTTCT CAACGGTCTT TTTTAAAGC ACATCTCAGA 1860
 20 TTAACAAAAA AAAAAAAA AAAAAAAA AAAAAAGGG CGGCCGC 1907

25
 (2) INFORMATION FOR SEQ ID NO: 109;

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 611 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

40 ATGAATTAAC GCCAAGCTNT NAATAGGGAC TCACTATGGG GGAAAGNTGG GTAACGCCCTG 60
 CAGGTACCGT TCCGGAATTC CCGGTCGAC CCACGCGTCC GATGGGGCTT TAGTAAATCA 120
 45 GCCTTGCAGG CTCAAAGCTG CAATCTGCCC ACTCTCAGGT ACTGAGACTT TGTGGGCCTC 180
 AGACACCAGG AAGAAAGTTG GGATACAGTC ATTTGAGTTA AAAAGGGAAT GACCCCTCAG 240
 AAACCCGCAT TAGCAGTGT ACTCTTGGA GTGCCTTTAC TTTTAACGCT CTCTGTTCTG 300
 50 AAAAGAGGT GTTTGGTTAC GTGTGAGCCA ACATCACGTT TTGTTAGCTG TGATTTACCT 360
 TTGTCCGTTT AAAAGACTTC ACGGAGCCAT TCTGTATACA AGGTGTGCTC TTCCCAATGT 420
 AGAAGGGGTT ATGGAAAAGG GTGGGATCCT TTGCTGTAAA CTGGAGAGAC CAGTCCCCAA 480
 CAGAGGGGAA TTTTAAGCCC TTCTCATCAC CCAATTGGAT GTTTTTGCTT ATAGCAAATT 540
 55 CCTGCAAAAT AAATAAATAA ATATTGCAA AACTAAAAA AAAAAAAA AAAAAAAA 600
 GGGGGNCCN C 611

60

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2632 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

10 TCCGAGCTCT CAGGACAAGG GCCTGGGCG ATCTTTTAA AAAGCCGATT GGGTGTCTTT 60
CTAAAANTAC AACCAGTACT TCATCGTCAA GTTTCCTGGA AGGGAGTCCC CTCCAGATTC 120
15 TCATGGAGTG ACAAATCTTG ACTCTTGCTC CTGGAATTTT TCAGGCCCAA ACTAGCGTTT 180
CTACAATGAT TTATTTGGCA AATTTGTCTT GATTATGGGT GGCTGATGAG GAACGTGCTT 240
20 TTGTTAGGAA CCGAAACTGG GCGCGGTGA GGGCGGTAC GCAATGAGTC CGGAAGAGGG 300
TGAAATGCTT TCGGTAGGCA CTCCACGGCT GTGAAGATGG CCGCGGCTGC GTGGCTTCAG 360
GTGTGCGCTG TCATTCTTCT GCTTCTGGA GCTCACCCT CACCACTGTC GTTTTTCAGT 420
25 GCGGGACCGG CAACCGTAGC TGCTGCCGAC CGGTCCAAAT GGCACATTCC GATACCGTCG 480
GGGAAAAATT ATTTTAGTTT TGGAAAGATC CTCTTCAGAA ATACCACTAT CTTCTGAAG 540
TTTGATGGAG AACCTTGTA CCTGTCTTTG AATATAACCT GGTATCTGAA AAGCGCTGAT 600
30 TGTTACAATG AAATCTATAA CTTCAAGGCA GAAGAACTAG AGTTGTATTT GGA AAAACTT 660
AAGGAAAAAA GAGGCTTGTC TGGGAAATAT CAAACATCAT CAAAATTGTT CCAGAACTGC 720
35 AGTGAAGTCT TTAAACACA GACCTTTTCT GGAGATTTTA TGCATCGACT GCCTCTTTTA 780
GGAGAAAAAC AGGAGGCTAA GGAGAATGGA ACAAACCTTA CCTTTATTGG AGACAAAACC 840
GCAATGCATG AACCATGCA AACTTGGCAA GATGCACCAT ACATTTTAT TGTACATATT 900
40 GGCATTTTAT CTTCAAAGGA ATCATCAAAA GAAATTCAC TGAGTAATCT TTTTACCATG 960
ACTGTTGAAG TGAAGGCTCC CTATGAATAC CTCACACTTG AAGACTATCC CTTGATGATT 1020
45 TTTTTCATGG TGATGTGTAT TGTATATGTC CTGTTTGGTG TTCTGTGGCT GGCATGGTCT 1080
GCCTGCTACT GGAGAGATCT CTTGAGAATT CAGTTTTGGA TTGGTGCTGT CATCTTCCTG 1140
GGAATGCTTG AGAAAGCTGT CTTCTATGCG GAATTCAGA ATATCCGATA CAAAGGAAAA 1200
50 TCTGTCCAGG GTGCTTTGAT CTTGTCAGAR CTGCTTTCAG CAGTGAAACG CTCACTGGCT 1260
CGAACCTTGG TCATCATAGT CAGTCTGGGA TATGGCATCG TCAAGCCACG CCTGGAGTCA 1320
55 CTCTTCATAA GGTGTAGTA GCAGRAGCCC TCTATCTTTT GTTCTCTGGC ATGGAAGGGG 1380
TCCTCAGAGT TACTGGGGCC CAGACTGATC TTGCTTCCTT GGCTTTTATC CCCTTGGCTT 1440
60 TCCTAGACAC TGCTTGTGTC TGGTGGATAT TTATTAGCCT GACTCAACA ATGAAGCTAT 1500

362

TAAAACTTCG GAGGAACATT GTAAAACTCT CTTTGTATCG GCATTTCCACC AACACGCTTA 1560
 TTTTGGCAGT GGCAGCATCC ATTGTGTTTA TCATCTGGAC AACCATGAAG TTCAGAATAG 1620
 5 TGACATGTCA GTCCGACTGG CGGGAGCTGT GGGTAGACGA TGCCATCTGG CGCTTGCTGT 1680
 TCTCCATGAT CCTCTTTGTC ATCATGGTTC TCTGCCGACC ATCTGCAAAC AACCAGAGGT 1740
 10 TTGCCTTTTC ACCATTGTCT GAGGAAGAGG AGGAGGATGA ACAAAGGAG CCTATGCTGA 1800
 AAGAAAGCTT TGAAGGAATG AAAATGAGAA GTACCAAACA AGAACCCTAT GGAAATAGTA 1860
 AAGTTAACAA AGCAGAGGAA GATGATTTGA AGTGGGTAGA AGAGAATGTT CCTTCTTCTG 1920
 15 TGACAGATGT AGCACTTCCA GCGCTTCTGG ATTCAGATGA GGAACGAATG ATCAGACACT 1980
 TTGAAAGGTC CAAATGGAG TAAGGAATGG GAAGATTTGC AGTTAAAGAT GGCTACCATC 2040
 20 ACGGAAGAGA TCAGCATCTG TGTCACTCTT CTGTACGGCT CCATGGGATT AAAGGAAGCA 2100
 ATGACATCCT GATCTGTTCC TTGATCTTTG GGCATTCGAG TTGGCGAGAG GTGTCAGAAC 2160
 AAAGAGAACA TCTTACTGAA AACAAGTTCA TAAGATGAGA AAAATCTACG AGCTTCTTAT 2220
 25 TTACAACACT GCTGCCCCCT TTCCTCCCAG ACTCTGACAT GGATGTTTAT GCAACTTAAG 2280
 TGTGTTGTTT CTGAACTTTC TGTAATGTTT CATTTTTTAA ATCTGACAAA CTAATAAGTT 2340
 30 TAACGTCTTC TAAAGATTG TCATCAACAC CATAATATGT AATCTCCAGG AGCAACTGCC 2400
 TGTAATTTTT ATTTATTTAG GGAGTTACAT AGGTGATGGG GGAAATTTGT AACTACCTTT 2460
 CATTTTCCTG GGAAGTCAAG GTTACATCTT GCAGAGTTG TTTTGAGAAA AAAGGGCCCT 2520
 35 TCTGACTTAA GGAGCCATAG TTCTATCAAT GATCAAAAGA AAAAAAAAAA AACTCGATCG 2580
 GCACGAGGGG GGGCCCGTA CCCAATTCCG CCTATGGGAN TCGAATGAGA CC 2632
 40

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2249 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

GAATTCGGCA CGAGCTCACC GTGCTGCGTG ACACAAGGCC AGCCTGCGCC TACGAGCCCA 60
 TGGACTTTKT RATGGCCCTC ATCTACGACA TGGTACTGSW TGTGGTCACC CTGGGGCTGG 120
 55 CCTCTTTCAC TCTGTGCGGC AAGTTCAAGA GGTGGAAGCT GAACGGGGCC TTCTCTCTCA 180
 TCACAGCCTT CCTCTCTGTG CTCACTGGG TGGCCTGGAT GACCATGTAC CTCTTCGGCA 240
 60 ATGTCAAGCT GCACGAGGGG GATGCCTGGA ACGACCCAC CTTGGCCATC ACGCTGGCGG 300

	CCAGCGCTGG	GTCTTCGTCA	TCTTCCACGC	CATCCCTGAG	ATCCACTGCA	CCCTTCTGCC	360
5	AGCCCTGCAG	GAGAACAACG	CCAACTACTT	CGACACGTGG	CAGCCCAGGA	TGCGGGAGAC	420
	GGCCTTCGAG	GAGGACGTGC	AGCTGCCGCG	GGCCTATATG	GAGAACAAGG	CCTTCTCCAT	480
	GGATGAACAC	AATGCAGCTC	TCCGAACAGC	AGGATTTCCT	AACGGCAGCT	TGGGAAAAG	540
10	ACCCAGTGGC	AGCTTGGGGA	AAAGACCCAG	CGCTCCCTTT	AGAAGCAACG	TGTATCAGCC	600
	AACTGAGATG	GCCGTCTGTC	TCAACGGTGG	GACCATCCCA	ACTGCTCCGC	CAAGTCACAC	660
15	AGGAAGAMAC	CTTTGGTGAA	AGACTTTAAG	TTCCAGAGAA	TCAGAATTTT	TCTTACCGAT	720
	TTGCCCTCCCT	GGCTGTGTCT	TTCTTGAGGG	AGAAATCGGT	AACAGTTGCC	GAACCAGGCC	780
	GCCTCAGACC	CAGGAAATTT	GGAAATCCTA	GCCAAGGGGA	TTTGGTGTA	ATGTGAACAC	840
20	TGACGAACTG	AAAAGCTAAC	ACCGACTGCC	CGCCCTCCCT	CTGCCACACA	CACAGACAGC	900
	TAATACCAGA	CCAACCTCAA	TCCCGCCAAA	CTAAAGCAAA	GCTAATTGCA	AATAGTATTA	960
25	GGCTCACTGG	AAAATGTGGC	TGGGAAGACT	GTTTCATCCT	CTGGGGGTAG	AACAGAACCA	1020
	AATTCACAGC	TGGTGGGGCA	GACTGGTGTT	GTTTGGAGGT	GGGGGGCTCC	CACTCTTATC	1080
	ACCTCTCCCC	AGCAAGTGCT	GGACCCAGG	TAGCCTCTTG	GAGATGACCG	TTCCGTTGAG	1140
30	GACAAATGGG	GACTTTGGCA	CGGGCTTTGC	CTGGTGGTTT	GCACATTTCA	GGGGGGTCAG	1200
	GAGAGTTAAG	GAGGTTGTGG	GTGGGATTCC	AAGGTGAGGC	CCAACTGAAT	CCTGGGGTGA	1260
35	GCTTTATAGC	CACTAGAGGT	GGAGGGACCC	TGGCATGTGC	CAAAGAAGAG	GGCCTCTGGG	1320
	TGATGAAGTG	ACCATCACAT	TTGGAAAGTG	ATCAACCACT	GTTGCTTCTA	TGGGGCTCTT	1380
	GCTCTAGTGT	CTATGGTGAG	AACACAGGCC	CGCCCTCTTC	CCTTGTAGAG	CCATAGAAAT	1440
40	ATTCTGGCTT	GGGGCAGCAG	TCCCTTCTTC	CCTTGATCAT	CTCGCCCTGT	TCCTACACTT	1500
	ACGGGTGTAT	CTCCAAATCC	TCTCCCAATT	TTATTCCCTT	ATTGATTTCA	AGAGCTCCAA	1560
45	TGGGGTCTCC	AGCTGAAANS	CCCTCCGGGA	GGCAGGTTGG	AAGGCAGGCA	CCACGGCAGG	1620
	TTTTCCGCGA	TGATGTCACC	TAGCAGGGCT	TCAGGGGTTT	CCACTAGGAT	GCACAGATGA	1680
	CCTCTCGCTG	CCTCACAAGC	AGTGACACCT	CGGGTCCTTT	CCGTTGCTAT	GGTGAAAATT	1740
50	CCTGGATGGA	ATGGATCACA	TGAGGGTTTC	TTGTTGCTTT	TGGAGGGTGT	GGGGGATATT	1800
	TTGTTTTTGGT	TTTTCTGCAG	GTTCCATGAA	AACAGCCCTT	TTCCAAGCCC	ATTGTTTCTG	1860
55	TCATGGTTTC	CATCTGTCTT	GAGCAAGTCA	TTCTTTGTTT	ATTTAGCATT	TCCAACATCT	1920
	CGGCCATTCA	AAGCCCCCAT	GTTCTCTGCA	CTGTTTGGCC	AGCATAACCT	CTAGCATCGA	1980
	TTCAAAGCAG	AGTTTAAACC	TGACGGCATG	GAATGTATAA	ATGAGGGTGG	GTCCCTCTGC	2040
60	AGATACTCTA	ATCACTACAT	TGCTTTTTCT	ATAAACTAC	CCATAAGCCT	TTAACCTTTA	2100

AAGAAAAATG AAAAAGGTTA GTTTTTCGGG GTCGGGGGAG GACTGACCGC TTCATAAGCC 2160
 AGTACCTCTG AGGTGAGTAT GTTTCATTA ACCTTTTGAT ATTTCTCRAA AAAAAAAAAA 2220
 AAAAAACCCG GGGGGGGGGG CGGACTCG 2249

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1198 base pairs
- (B) TYPE: nucleic acid
- (C) STRAININESS: double
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GATACTATTA GGCAGTGC TCAGCGGTGC GCGGTAGAC TAGTGATCC CCGGTGCAGG 60
 AATTCCGCG AGCGCGCGCG GAGTCGAGT GTGGCGCGCC CCGCGCGCGC TGCCTCCGCG 120
 GANCCCAAAA TCATGAAGT CAGCTGAG AGCCCGAAGA AAAGGAGGAA TTCGCCGTGC 180
 CCGAGAATAG CTCCGTCCAG CAGTTAAGG AAGAAATCTC TAAACGTTTT AAATCACATA 240
 CTGACCAACT TGTGTTGATA TTTGTTGGAA AAATTTTGAA AGATCAAGAT ACCTTGAGTC 300
 AGCATGGAT TCATGATGA CTTACTGTT AGCTTGTCAT TAAAACACAA AACAGGCCCTC 360
 AGGATCATTC AGTCAGGA ACAATAGAG CTGGAAGCAA TGTACTACA TCATCAACTC 420
 CTAATAGTAA CTCTACATCT GGTCTGCTA CTAGCAACCC TTTTGTTTTA GGTGGCCTTG 480
 GGGGACTTGC AGGTCTGAT AGCTTGGGTT TGAATACTAC CAACTTCTCT GAAGTACAGA 540
 GTCAGATGCA GGCACACTT TGTCTAACC CTGAAATGAT GGTCCAGATC ATGGAAAAC 600
 CCYTGTGTTCA GACCATGCTC ATCAATCCCT GACCTGATGN AGACAGTTAA TTATGGCCAA 660
 TCCACAAATG CAGCAGTTGA TACAGAGAA TCCCAGAAAT TAGTCATATG TTGAATAATC 720
 CAGATATAT GAGACAAAG TTGCACTTG CCCAGGAATC CAGCAATGAT GCAGGAGATG 780
 ATGAGGAACC AGCAACCGGC TTGAGCAAC CTAGAAAGCA TCCCAGGGGG ATATAATGCT 840
 TTAAGGGCCA TGTACACGA TATTGAGGA CCAATGCTGA GTGCTGCACA AGAGCAGTTT 900
 GGTGTAATC CATTGTGCTC CTTGCTGAGC AATACATCCT CTGGTGAAGG TAGTCAACCT 960
 TCCGTACAG AAAATAGAA TCCCTACCC AATCCATGGG CTCCACAGAC TTCCAGAGT 1020
 TCATCAGCTT CCAGCGGCAC TGCTAGCACT GTGGGTGGCA CTACTGGTAG TACTGCCAGT 1080
 GGCATTCTG GGCAGAGTAC TACTGGCCA AATTTGGTGC CTGGAGTAGG AGCTAGTATG 1140
 TTCAGACAC CAGCAATGCA GAGTTGTTG CACCAATAA CTGAAAACCC ACAACTTATG 1200

365

CAAACATGT TGTCTGCCCC CTACATGAGA AGCATGATGC AGTCACTAAG CCAGAATCCT 1260
 GACCTTGCTG CACAGATGAT GCTGAATAAT CCCCTATTTC CTGGAAATCC TCAGCTTCAA 1320
 5 GAACAAATGA GACAACAGCT CCCAACTTTC CTCCACAAA TGCAGAATCC TGATACACTA 1380
 TCAGCAATGT CAAACCCCTAG AGCAATGCAG GCCTTGTTAC AGATTGAGCA GGGTTTACAG 1440
 10 ACATTAGCAA CGGAAGCCCC GGGCCTCATC CCAGGGTTTA CTCTGGCTT GGGGGCATT 1500
 GGAAGCACTG GAGGCTCTTC GGGAACTAAT GGATCTAAGC CCACACCTAG TGAAAACACA 1560
 AGTCCCACAG CAGGAACCAC TGAACCTGGA CATCAGCAGT TTATTGAGCA GATGCTGCAG 1620
 15 GCTCTTGCTG GAGTAAATCC TCAGCTACAG AATCCAGAAG TCAGATTTCA GCAACAACCTG 1680
 GAACAACCTCA GTGCAATGGG ATTTTGAAC CGTGAAGCAA ACTTGCAAGC TCTAATAGCA 1740
 ACAGGAGCTG ATATCAATGC AGCTATTGAA AGGTACTGG GCTCCAGCC ATCATAGCAG 1800
 20 CATTTCTGTA TCTKGAAAAA ATGTAATTTA TTTTGTATAA CGGCTCTTAA ACTTTAAAAT 1860
 ACCTGCTTTA TTTCATTTTG ACTCTGGAA TTCTGTGCTG TTATAAACAA ACCCAATATG 1920
 25 ATGCATTTTA AGGTGGAGTA CAGTAAGATG TGTGGTTTTT TCTGTATTTT TCTTTTCTGG 1980
 AACAGTGGGA ATTAAGGCTA CTGCATGCAT CACTTCTGCA TTTATTGTAA TTTTTAAAA 2040
 ACATCACCTT TTATAGTTGG GTGACCAGAT TTTGTCTGC ATCTGTCCAG TTTATTGCT 2100
 30 TTTTAAACAT TAGCCTATGG TAGTAATTTA TGTAGAATAA AAGCATTAAA AAGAAGCAA 2160
 AAAAAAAAAA AAAAATTCCT GCGCCCGCA ATTCTTCT 2198

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1043 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

CTGAAGTGTA TGTGGTGAGG AAGAAGAGGC TCCTACTGTA GACAGCCTTG TTCTACAGAT 60
 50 CCTCCCAGAA ATCTCTGGGC CAGGTGGAAC CCAGGGTCAG AGAGGGATGG GAGAGAGGTT 120
 TAATTTTCCA TGATAAATAA AAATCTATAA AATAATAAAC AAGAGAAAAG AGATTGGAAG 180
 55 CAGCCAGGTT GGAGCAGTGA GTGAGTAAGC AAACCTGGCT GGCCTCTCCA GATTCCCCAG 240
 GCTCTCAGAG AAGATCAGCA GAAAGTCTGC AAGACCCTAA GAACCATCAG CCCTCAGCTG 300
 CACCTCTCTC CTCCAAGGA TGACAAAGGC GCTACTCATC TATTTGGTCA GCAGCTTTCT 360
 60 TGCCCTAAAT CAGGCCAGCC TCATCAGTGG CTGTGACTTG GCCCAGGTGC TGCAGCTGGA 420

366

5 RGACTTGGAT GGGTTTGAGG GTTACTCCCT GAGTGACTGG CTGTGCCTGG CTTTTGTGGA 480
 AAGCAGTTC AACATATCAA AGATWAATGA AAATGCAGAT GGAAGCTTTG ACTATGGSCT 540
 CTTCCAGATC AACAGCCACT ACTGGTGCAA CRATTATAAG AGTTACTCGG AAAACCTTTG 600
 CCACGTAGAC TGTCAAGATC TGCTGAATCC CACCTTCTT GCAGGCATCC ACTGCCCAA 660
 10 AAGGATTGTG TCCGGAGCAC GGGGATGAA CAACTGGGT AGAATGGAAG KTTGCACTGT 720
 TCAGGCCGGC CACTCTTCTA CTGGCTGACA GGATGCCGCC TGAGATKAAA CARGGTCCGG 780
 GTGCACCGTG GARTCATTC AAGACTCCTG TCCTCACTCA RGGATTCTTC ATTTCTTCTT 840
 15 CCTACTGCCT CCACTTCATG TTAATTTCTT CCCTTCCCAT TTACAACATA AACTGACCAG 900
 AGCCCCAGGA ATAAATGGTT TTCTTGGCTT CCTCCTTACT CCCATCTGGA CCCAGTCCCC 960
 20 TGGTTCTGT CTGTTATTTG TAAACTGAGG ACCACAATAA AGAAATCTTT ATATTTATCG 1020
 AAAAAAAAAA AAAAAAACT CGA 1043

25

(2) INFORMATION FOR SEQ ID NO: 114:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 703 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

40

45

50

55

60

GAATTCCGCA CGAGTGCCCG GGCACCACGG CGGTTTTTCG ACGCTGCCCG TGGACCCAGG 60
 CAGCATGGAC CACGGTTGCT GGGCGGATGG GGAGCGTCTA TGGTCAGTTG CCTTAGAAGT 120
 GGTGAGATCG GAAGCTGCAG TTGGAAGACC CTGCAGGATG CCTGACAACG GGATGTCTGA 180
 CACATGATTG GAGCTCTTTT TGAAATGTTT CTGCCCCTTC CTGGAGCAGA GGAGCCATTA 240
 TTTATGCAGG TACATCGAAG TCTTTTGACC TCATACAGT GATTATGCTT GTCATCGCTG 300
 GTGGTATCCT GCGCGCCTTG CTCCTGCTGA TAGTTGTCGT GCTCTGTCTT TACTTCAAAA 360
 TACACAACCG GCTAAAAGCT GCAAAGGAAC CTGAAGCTGT GGCTGTAAAA AATCACAACC 420
 CAGACAAGGT GTGGTGGGCC AAGAACAGCC AGGCCAAAAC CATTGCCACG GAGTCTTGTC 480
 CTGCCCTGCA GTGCTGTGAA GGATATAGAA TGTGTGCCAG TTTTGATTCC CTGCCACCTT 540
 GCTGTTGCCA CATAAATGAG GGCCTCTGAG TTAGGAAAGG TGGGCACAAA AATCTTCATG 600
 AGCAATACTT CTTAGTAGAT TGTTTTGTTA TTCAAATCAA GTTCTAGTGT TTTTATGTGA 660
 GATTATATAA TTTACAGTGT TGTTTTATAT ACTTTTGAAT AAA 703

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

5	GGCAGAGGGG GCATGAGCAG GAGGAGGATT ACCGCTACGA GGTGCTCAGG GCCGAGCAGA	60
10	TTCTACAACA CATGGTCGNA ATGTATCCGG GAGGTCAACG AGGTCAATCCA GAATCCAGCA	120
	ACTATCAGAA GAATACTCCT TAGCCACTTC AATTGGGATA AAGAGAAGCT AATGGAAAGG	180
15	TACTTTGATG GAAACCTGGA GAAGCTCTTT GCTGAGTGTC ATGTAATTAA TCCAAGTAAA	240
	AAGTCTCGAA CACGCCAGAT GAATACAAGG TCATCAGCAC AGGATATGCC TTGTCAGATC	300
20	TGCTACTTGA ACTACCCTAA CTCGTATTTT ACTGGCCTTG AATGTGGACA TAAGTTTTGT	360
25	ATGCAGTGCT GGAGTGAATA TTAACTACC AAAATAATGG AAGAAGGCAT GGGTCAGACT	420
	ATTTCTGTTC CTGCTCATGG TTGTGATATC TTAGTGGATG ACAACACAGT TATGGCCTG	480
30	ATCAGAGATT CAAAAGTTAA ATTAAAGTAT CAGCATTTAA TAACAAATAG CTTTGTAGAG	540
	TGCAATCGAC TGTTAAAGTG GTGTCTCGCC CCAGATTGCC ACCATGTTGT TAAAGTCCAA	600
35	TATCCTGATG CTAAACCTGT TCGCTGCAAA TGTGGCGGCC AATTTTGCTT TAACTGTGGA	660
	GAAAATTGGC ATGATCCTGT TAAATGTAAG TGGTTAAAGA AATGGATTAA AAAGTGTGAT	720
	GATGACAGTG AAACCTCCAA TTGGATTGCA GCCAACACAA AGGAATGTCC CAAATGCCAT	780
40	GTCACAATTG AGAAGGATGG TGGTTGTAAT CACATGGTCT GTCGTAACCA GAATTGTAAA	840
	GCAGAGTTTT GCTGGGTGTG TCTTGGCCCA TGGGAACCAC ATGGATCTGC CTGGTACAAC	900
45	TGTAACCGCT ATAATGAGGA TGATGCAAAG GCAGCAAGAG ATGCACAGGA GCGATCTAGG	960
	GCAGCCCTGC AGAGGTACCT GTTCTACTGT AATCGCTATA TGAACCACAT GCAGAGCCTG	1020
	CGCTTTGAGC ACAAACCTATA TGCTCAGGTG AAACAGAAAA TGGAGGAGAT GCAGCAGCAC	1080
50	AACATGTCTT GGATTGAGGT GCAGTCTCTG AAGAAGGCAG TTGATGTCTT CTGCCAGTGT	1140
	CGTGCCACAC TCATGTACAC TTATGTCTTC GCTTTCTACC TCAAAAAGAA TAACCAGTCC	1200
55	ATTATCTTTG AGAATAACCA AGCAGATCTA GAGAATGCCA CAGAGGTGCT CTCGGGCTAC	1260
	CTTGAACGAG ATATTTCCCA AGATTCTCTG CAGGATATAA AGCAGAAAGT ACAAGACAAG	1320
	TACAGATACT GTGAGAGTCG ACCAAGCGTT TTGTTACAGC ATGTGCATGA AGGCTATGAA	1380
60	AAAGATCTGT GCGAGTACAT TGAGGCTGTA GAATGGCCCT GCATAAAAATG AACTCTGAAA	1440

	ACTTTACCAT CTAGAGTGCT CATGCAATTA AAACAAAACA AACACAAACA AGGAGGCTCT	1500
5	AAGCCTATTC TGACACCACT GGTCTGTAGT ACCAGAATTG TTTTGTATAT GGAAAGTTTA	1560
	AGTAAATTAT ATTGTAATAA AAAGGTAGAT AAACCATTTG ACACACATAT TCTAGGCCCC	1620
	CAACAAAAGT GTGACAGACA CACTAAAAGC CCTCCAACTT TAACTTGTA CGTAGCTTCA	1680
10	TTCTCAAAGC TGACTCCTTT TTTTCTTTT TCCTTTTCCT GAGTGTAGTA CAGTTAAAT	1740
	TTCAAACAGC TCCTTGACAC TGCTTTTCAT GTTCAAACCA GCCATTTTGT TGTACTTTGG	1800
15	TAAAGGACCT CTTCCTCTTC CTCCCTACA CATACAGATA CACCCACACA CAGACTGACT	1860
	CTCTTTCTCT CATACCCCA GGTGATGAGT GAATGATGCT TAGTTCTTG TAAAGAAAAT	1920
	CTTGGGATGG GGAAAGGGT AGGCAGCAAG AGGATTCAAC AAACGAAAAA CATAAAAACT	1980
20	TTGTATATGA CTTTAAAAAC AAGAGGACAA CACAGTATTT TTCAAAATTG TATATAGCCG	2040
	ATATGCATGG ACAAGCAAG CGTGGCACGT GTTTGCATAA TGTTTAATTA CAAAAAATA	2100
25	TTTATCTTTT AAAAATCTTC AAGATTATGT CTATTTGCTG TGCATTTTCT TTCAGTTTGC	2160
	TTATCTTTCC CGGTTTGGG TTGGGATAAA GGTGTGTCCG TTTAGCACCT CTGGAAGACC	2220
	TATCTAGAGC TCTTTCACTT TCCTGACGTT ATTTTGCCCT TTCTGGTGTG GGTATGTCTG	2280
30	TTGCCGCCCA TGGGCTNCAY GCCTTGAATT CCTGCTCTTG ATCAGGGACA AGGGAGGTCA	2340
	AGCTCTGACT AATGCCATGA CCGATTAAAG GGGTACAGCA GGGAGTTTGT TTGCTACAGC	2400
35	TCATGAATTA ACCTGTCCCA ACCTAATCCC CCTCCATGGC ATCATGCCCT TACCCAAGCC	2460
	TTTGTGTGCC CATGTTATGC ACACAGCTGT AGGCATTCTT AAGTCCCTG TCGCATCCAG	2520
	TGGAAGCATT TTAATAATTC TTTTACTTTT TGGTTTCCC TTAATTGCTG CTTTTCAGAT	2580
40	TTTAGTTATG GCTCGTCTGC TCACCCCTTC TCTACATTAG GGTGTCAAAG AGAATGTTTT	2640
	GCTTTAAATA TAAATAGCCA TTCATTTAGT CTCAGATTGT GAATTTAAAA TGGTGGATAC	2700
45	CGAAATTGCT TGTGTGTGTT GCTGTGGGTT TGGTTTGAAG GCAAACACCC CTAGAACATG	2760
	ATATTTCCCAT CTAGTGCAAT TAAATAGAAA TCACTGAGTT TGCTGCTTTT TTATTGTCTG	2820
	CAGATAGGAG AATTAATAAT GCATTTTAGC TGTGATGTCC ATTTTATGA AATTCCTACT	2880
50	AAGAGCTATG TTAAGTAAGT AGGATGGTGG TGGTTGTATT AACTATATAC CTGTTTAGGC	2940
	CATCTGGCT GTGGTATTTT TCAATAGGTC AGCATCTGTA AATCTGTCTG TTTTATACAG	3000
55	GAGTGCAGAG TGAAC TAGGC AACTAGATTA AGAGGTCTAA ATATGAAATA CCAGTTGAGG	3060
	CTGAGGACCT CTTGCTCTTC CTTTAAATGT CTTTTCCTA GGGAGTGTTC ACCATTGTG	3120
	AGGCAGCTTT GTCTGCTCTT AACTGTACA TCCTATTACT CCATTGGGAA GTAGGTTTAC	3180
60	TTTCTCTGG CTTTTCCTT AAGTTAGGCT TTGCTGAATC AACCTTACTT TCTCTTTAG	3240

5 AAAAGGTTGT TACAGGAGAT TTAAGTGGCA CTGTTCTTTT CCCATCAAAA ATCAGTGAAT 3300
GTTTGCTGAG TATAAATGCT GCTTCCTTAA ACCACTTGTC GCTTTAGGAT CAACTTTACC 3360
TGTACCTTTT CTCCTTTCTT CCCTTGCCAC CTCAGGTGCA AATCTGAACT CAGTGTCTGC 3420
TTCTTCCATT TTCTCGTCTC TCTCCCTCTT TCCCCCAFTA TCCATATGAC ATTATTTTAC 3480
10 TTCAAATGAC AGCATCAATC TTAATAAGAT ATACATTAAA ACTAAGGAGT TTTTAAAG 3540
AAAGCCTGAA TAAGTTCTT TCCCTGGTAA CTTTGAAAAG CAGTCAGAGT TCCTATATAG 3600
ATATATGTGG CTCCTTTAAA ATGCTTTGTG TATGTGTGGT GTTTAAAAAA AAAAAAAAAA 3660
15 TTGGGGGGGG GGCCCGGTTC CCAT 3684

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1965 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

30 AAGAAAGGGT ATTAAAATTC TAGATCACAT ATGGACCCGG GAAGGTTTTT NACCCTCTGT 60
TAGTGACATC GAGTCTCCCA CTAGACAAAA TAGGTGGAAA AATCTCTCGA GGGCTCACAT 120
35 TGTTTTGTCA TCTTCAGGAA AAACACCACC AGGCCATACC ACAGCCTGCC CAGTGAGGCG 180
GTCTTTGCCA ACAGCACC GGATGCTGGT GTGGCCTTTG GGCTGCTGGT GCTCTACATC 240
40 CTTCTGGCTT CATCTTGGA GCGCCAGAG CCGGGGATCC TCACCGACAG ACAGCCCTCG 300
CTGCATGATG GGGAGTGAAG CAGCAGGAAG GGGCTCCCAA GAGCTCCTGG TGSTGCAGCC 360
TGTGCTCCCC TCAGAAGCTC TGCTCTTCCC AGGGCTCCCC GCTGGTTTCA GCAGGCGACT 420
45 TTCTTCCAAT GCTGGGCCCC GACTTCTTGC CTGGGTGCTG GCCTGCCCTC TCCGGNCCGC 480
TTGCTGCCTG TCTGCTTCC TTGGTGGYTT TGCTGGGTGC TGGGCCTGCC CTCTCCGGCC 540
GCTTGCTGCC TGTCTGCTTT CTTTGGTGGC TTTGCTGGGT GCTGGGCTGC CCTTCTCTGG 600
50 CTGCTTGCTG CCGTCTGCTT TTCTTGGTG GCTTTGGCTT CTGCACTCCT TGGCGTCASC 660
TCTCAGGTCC TCCATTCACA CGAGTCTCTC CTCGCTCTGG CCGCTCTTGC TGCTCTGTG 720
55 TGAAGAWATC AGACTGATTT CCTCTTAAGA CTCCTAGGGA TGTGGTGAAG AGCTGGGACT 780
CAAGTGCACT CCACGGTGTG AAACATGAGG GARGTGAGGT GTCCGTCCAC TTCCCCCATA 840
AAGGTGTGCA TTTCACTTAG GCTGCCCCGC CACAGAGCAG GCTTCATCTG CTCTGCCATC 900
60

370

CAGCCCCATC TGGATGTGAG GTGGGGTGA GACATCATGG GGTGATTGCA GAAAGGGGGA 960
CTGGCGGCCCC ACCGAGCTTC TGCTGAGSAG CTGACCGCTC TGAGCTGTTC TGTTCGTAT 1020
5 TGCTGCTCTG TGTCTGCATG TATTGTGACC GTGGGGCTCC ACCTCTTCCA GCTGCTGCTA 1080
CAGCTGAGGC CTGGATCCCC GCCTTTCCCT GTGACTTACG TGTCTGTAC CCGCANGCAG 1140
10 CCTACCAAT CCTGGTGACC TGCTCTCCCA AGAACAGAGC CTGTCCCCAG ATGTCCCACT 1200
AGCGATGAGT AACAGAGGTG GCTGTGGACT TCCTCTACTT CTCCTTGCTG GATCAGGGCC 1260
TTCTTGCTTC CCGCTGGGCA GGTCTGGCCT TGCTCTCTTG GCAGGGCCCC AGCCCTCTG 1320
15 ACCACTGTGC AGCTCACCAT GCAGCTGATG CCAAGTTGT GGTGTCCACT GTGCAGCAGC 1380
CCTGGGAGCC ACTGCCACCT TCAGAGGGGT TCCTTGCTGA GACCCACATT GCTTCACCTG 1440
GCCCCACCAT GGCTGCTTGC CTGGCCCAAC CTAGCGTTCT GTGCCATGCT AGAGCTTGAG 1500
20 CTGTGCTCT TCTTCAGCGG AGGAAATAGG GTGCAGAGCG GGAAGGTCT TGCTCTAAG 1560
TGTTGCTGCT GTGGCTTTT TCCCTCTCC AAAGACGCAC TGCCAGGTCC CAAGCTTCAG 1620
25 ACTGCTGTGC TTAGTAAGCA AGTGAGAAGC CTGGGGTTTG GAGCCACCT ACTCTCTGGC 1680
AGCATCAGCA TCCTACTCCT GGCAACATCA GGCCAACGTC CACCCAGCC TCACATTGCC 1740
AGATGTTGGC AGAAGGGCTA ATATTGACCG TCTTGACTGG CTGGAGCCTT CAAAGCCACT 1800
30 GGGATGTCT CCAGGCACCT GGTCCCATG ACCAGCTCCC CGTCTCCATA GGGTAGGCA 1860
TTTCACTGGT TTATGAAGCT CGAGTTTCAT TAAATATGTT AAGAATCAA GCTGTCTTTG 1920
35 TTCAGGCTGC TATAACAAA ATATAATAGC CTGGGTGGCT TAAAC 1965

40 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 503 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

50 AGTGATCCCC TTGCTCGGC CTCCTAAAAT GCTGGAATG TAAGCGTGGG CCTCTGCACC 60
CGGCTGTGTC CGCAATTAA AAACGCACAG CCACCATCC CTCTCCAGAA AGCACCAGCA 120
TGCCTTTGGG AGAACCAGCC TCCTCCATGG AGGAAAGCTT GGGATCTGCC TTCCACCTG 180
55 GGGAGGAGAG GATCTGTGG AAAATCCTTC TGACGGACTT CCCCTCAGTG CCTCATCCAT 240
ACTCAATAGT AGAAAAAGTA AGAAATATAC AAAGATAGCA GATACACGGA GACAGTTCCC 300
60 CAAATAGCTG ACCGAWTAGC GCAGAAGCAA TATTGAAGAC CTAATAGCTG AGACATTTCC 360

371

AGAACTGATA AAGTGCATCC AGCCACAGAT CAAGCAGCCC AGAAAATTCC AGGCAGCATC 420
AACAAATAAA TAGCCCCACA TGCACCCGTG AAAATGCAGA AGACCAACA AAAAAGTCCG 480
GTCAACAGCC AGAGTTAAAG AGG 503

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1133 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

GGCAGAGCTT GGAATGAACC CCTGTGGATA AGGGGGACTA TTAGATAGAA TAAACATCAA 60
TAAATGCTTG ATGAATAAAC GCTAATCCTA CCTTCCAGC CTGACACCTC CCAGTGGACA 120
CCACACTTCA CTGGAAGCCT TAGAAACCTT TCCACCCAT CCTTCCAGCC CTGGCTTCAT 180
GTTGCCATTT CTCACCCCCA GAACAGGCCG CCCGCTGAA GAAACTACAA GACCAAGAGA 240
AACACAGAA AGTGGAGTTT CGTAAAAGGA TGGAGAAGCA GGTGTCAGAT TTCATTCAAG 300
ACAGTGGGCA GATCAAGAAA AAGTTTCAGC CAATGAACAA GATCGAGAGG AGCATACTAC 360
ATGATGTGGT GGAAGTGGCT GGCCTGACAT CCTTCTCCTT TGGGAAGAT GATGACTGTC 420
GCTATGTCAT GATCTTCAA AAGGAGTTTG CACCCTCAGA TGAAGAGCTA GACTCTTACC 480
GTCGTGGAGA GGAATGGGAC CCCAGAGAGG CTGAGGAGAA GCGGAACNTG AAGGAGCTGG 540
CCCAGAGGCA ANGAGGAGGA GGCAGCCAG CAGGGGCTG TGGTGGTGAG CCTGCCCAGC 600
GACTACAAGG ACAAGTACAG CCACCTCATC GGCAAGGGAG CAGCCAAAGA CGCAGCCCAC 660
ATGCTACAGG CCAATAAGAC CTACGGCTGT KTGCCCGTGG CCAATAAGAG GCACACAGCC 720
TCCATTGAAG AGGCTATGAA TGAGATCAGA GCCAAGAAGC GTCTGGGCA GAGTGGGGAA 780
GAGTGGCCGC CAACCTCCTA GGCGCCCCGC CCAGCTCCCT TTGACCCCTG GGGCAGGGCA 840
GGGGCCAGGG AGACACAAGG CTGCTGCTAT TAGAGCCCAT CCTGGAGCCC CACCTCTGAA 900
CCACCTCCTA CCAGCTGTCC CTCAGGCTGG GGGAAAACAG GTGTTTGATT TGTACCGTT 960
GGAGCTTGA TATGTGGTG CATGTGTGT GTGTGTGTGA GAGTGTGAAT GCACAGGTGG 1020
GTATTTAATC TGTATATTG CCCGTCTTG GAATTTTCTT CCCATGGGGC TGGGCTACTT 1080
TACATTCAAT AAATACTGTT TAACCCAAAA AAAAAAAAAA AAAAGAAAGA AGN 1133

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1101 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

GGGCACAGCT GAAGCTGCAG ACCTCCCCAG GGGATGGGTC CTCTCCCCCA GGAGCCCCGA 60
GGCAGGGGAG GCAGAAAGCC TGGGCTCTGG GGGGTGGCCT GGGACAGCT GTGCTGTGGG 120
CCGGGGGCTG GGCCTGTCCC ACAGCGNCCT GGAGCTCGTG GTTCTGAGCA GCCAGCTGGG 180
TGGTGTCTGG GGATAGCTGG GAGGCACAGC GGCTGCCATG TGGGACTGGG ACTGGAGTGC 240
TCCCTGGTCT TGGCCTCTGT GCTCAGCCT TGCTCTGGTC TGCTGAGTG CAGGGGCCAA 300
GGGGCACAGG GCCAGTGAGG CCGGCCACGC TCGGGCCCTC ACCTGTGAGA TGGGTTCGGA 360
ATTTKACACA GCCTANGGCT TGGTTCTTGG TKGTNGAMCG TGGACTYCTK AGAAGGGGAG 420
TGCTGGTCCT GAAAGGGCTG GTTGGAGACG AGCTGCTTTT CTCGCTGTTT TTCTCTTAGG 480
AGATTAAACA AAAACAGAAA GCACAAGACG AACTCAGTAG CAGACCCCAG ACTCTCCCT 540
TGCCAGACGT GGTTCAGAC GGGGAGACGC ACCTCGTCCA GAACGGGATT CAGCTGCTCA 600
ACGGGCATGC GCGGGGGGCC GTCCCAAACC TCGCAGGGCT CCAGCAGGCC AACCGGCACC 660
ACGGACTCCT GGGTGGGGCC CTGGCGAACT TGTGTGTGAT AGTTGGGTTT GCAGCCTTTG 720
CTTACACGGT CAAGTACGTG CTGAGGAGCA TCGCGCAGGA GTGAGGCCCA GGCCCCGAGA 780
CCCAAGCGGC CACTGAGGGC ACCGCGCACC AGAGCGTGAC CTCGGCAGGC TGGACACACT 840
GCCCAGCACA GCGAGACCCA CCAGGCTCCT AGCTTTAGCT TTTAAAAACC TGAAGGGGA 900
AGCAAAAACC AAAATGTGTG ACTGGGCTTT GGAGGAGACT GGAGCCTCAG CCTGTCTCTG 960
GCCACGGGCC GCTGGGGCTG GTGTGGGTGG GCCTTGCTG CTGGATTTGT AGCTTATCTT 1020
CCGTGTGTG TTTGGACCTG TTTTAGTAAA CCGTTTTTTC ATTTTAAAAA AAAAAAAAAA 1080
AAACTTTGGG GGGGGGCCCC N 1101

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

AGCTTCTCTG TCCACTCTTG AACTCTGGGS TCTCTTGGAA CTTTCTCTAC CCTCTCTAGC 50
 5 CTGAATATTC CTTCATGGA TTCCACTCAA CCAGACTTTG GATCTGTGCC TACTTAATCA 120
 ACCTTATCTT TGCAATATGT TCGGGCCAC CTTCCACTCC TTGGTTCTTG TTCTCCTTG 180
 GCCTAACTTG TCCCTTCTCC ACTTCACATC CCGGTGGGA CAGCACTCCT CCTTCCTGCC 240
 10 AACCTCCTC CGTCTCARAA AAAAAAAAAA AAAAAAAAAA TT 282

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2635 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

25 TAAGGGGGTG TGTGCTCACC TCCTCCTGAC CTTTACACT CTTGTCTGC CCAGACCAAC 50
 AGAGAGAGCT GTCCCTGAGA CCGCGGAGAG AAGCAGCTGC CGAAAGCTGC AGCCTTTCCG 120
 30 CACTCTGAGA CATGATCTT CCTCCTGCCA GGGGAGAGCC ACCCAGAGCC CATGTCCAGC 180
 CCACTTCCC TCAGCCCCCA GGGTTCTCTT CTGGCCCTC TGAGGATTCC CTAGGGCTGC 240
 CCGCAGAGG GGTTCGCCA AGCTCTGTTT TGAAGCCTGC AATGTGAAA AGTGAGAAGT 300
 35 CAGAGGGAAC AGGACAGGTG CAGCCGGGCT CTGAGGCCAC ACCTCACACC TCGCTGTGCC 360
 CCAACATCCC CTGAGCAGTG TGAGCTCATC TCACCAGATG AGAAGAGCCC CTGTGCATTT 420
 40 YTTTTGTTTG TTTGTTGCTG TTTCCCCCA CCACTCCAGT TCTCTCAGC AAAGCAAATT 480
 CCTTAACACC TTTGGTGGAG AATTCTTAC CCAGACTTGG GCTGTGATG CCTTCAGTG 540
 CGTGGTGAGT GCAGCGTGTG TCGGTGTGCC TGTGTGTGAA CCTGGGGGCC ATCCTGGTGG 600
 45 CCTGGGAGCG TGAGGAGAGG CCCCCTGTGT GCTGGGTGAG TGGTGGGTGT GGGTCAATG 660
 CAGTGAGGCT CTCTGGGTGA GGCTCCCAAC CTGGCAGTCC CCAGCCTCCC AGCATCTGTG 720
 50 AGCGTCTGTT GCACTTTACA GAAGAGCCTC ATCTGTCTG CCGCTCACTC TGCCCTGGAA 780
 TCAACATCTT CCGAGTCTTT CTGGGGGAA ATAGCAGAGC CCACTTAAC TCCATAAACT 840
 GCTTCCCATT CCGCAGCCCA GTTCTGATTG TTGAGGTGTC GCGTCTGTCC AGGTCCCCCA 900
 55 GTCCCTCTCT TCTCTGTCC TCTCTGTCT CTTCACTTCC CCACTCCAGC CCGGGCTCAG 960
 TTCAGGGAAA TGCTGTCCA YATCAGCCCT CTGCTCTCTG AGGCAGCCGC GCCTCTGACT 1020
 60 CGGAGCTACT TGAACCTTCT GCTCTTGCTA GGAFTGGAGT CTACCTATCT CTTCATTTG 1080

	TCCCAGCTGG AGTTCCTGGAA CTTTCCTCCT CGGGSTGGGG GTGGGGSTTG TTAAAGATCC	1140
5	TGGGGGGGCTT GCGGAAGGAA GGAGTTCAGA GGAGGGTGT CCGCTGTCTT CTTGATGTCA	1200
	CCCTCCGCTC CTGGGACACG TGCTCTCTCT GTCTCTGGGT CTTCTGGSTG TGAAGCTTTC	1260
	TGTGTCTTG TAAATATGTT TTAGGAAGAA AGCAAAAGG ACTGAATAG CTTTTCGTAG	1320
10	GATTGCAGGG GTCCAGCCTT GCCTGTTTCC GAAGCGGSCA CACTGCTTTT CCGTCCACTG	1380
	AGACTGGTCC CCTCAAAAGG TAGACAAAAC AGCAGCTTCC TGTGGAGTTG AAGGCGGGCC	1440
15	TCAAAGTGGC TTTTGTGTAG ACAAGGTAA GGTTCCTCA TGAGCAAGT TGAAGATCGG	1500
	TCCTTCTCA GTCCTTGAT TGTGACCTT GACCAAGGG CCTGCGCTCC AGCTCTCCA	1560
	GTGCCCTCTC CTCGATGCTT GCCTGCTTCC TGGCCGCACT CCGCTGGTTT AAGGAGGTAG	1620
20	CGGAATTAGG GCCATGCTGG AAGAAGCTTA ACCATGTGTT CAAGAAAGG TTTTTCCTT	1680
	GCTTGGTCTT GGAAGTCCC TTGGCTGCCC GAGGCTTCTT TGGCCGCTG GTGTGCGGG	1740
25	AGGTGGATGT CAGATCTGGT AGGTTGCAGC AGAGAAATA AATGTGCTT GAGGACCCAC	1800
	TCAGAGAGGG TCCAAGGTG ATGGAGAAGG AAGCATGGCC TGGGAGTTTG GAAGGAGGG	1860
	GTGGTGGGTG GCGGCATTT GACTGCCCCC TGTGTCTCA CAGGTGCGGG GTGTGACCT	1920
30	CYCTTCACTC CAGCCCGGCT GCCTTCAGCC TTCCATGAGC TTCAGCTGCT TGAAGTTCA	1980
	CTTTGGAGGG GGTGGGGTCC GTTGGCATCA ACACGGGGAC CCTGTGTTTC ACCAAAGCCC	2040
35	GAGCCCTCAG CCGCTGGGGA GAACAAATGG CTGAGCTTTC ATAGCTGGGG TGTTCGAGAG	2100
	GCTGCGGGCT GCGGGCAGTC CCAGGGGAGA GACACCCACG AAGGAGATTC AGACATCCCG	2160
	AGGAAGTTCC CAGCAGAGCA AACTGCTTTC CAGGCTGAG CCGCTTAAA CTGTGTGATG	2220
40	TGCAATAACT GAGCTTAGAG TTAGGAATTC TGTTCAGTC CTTGGAATTC CTTCTGTAGA	2280
	TTTAAGTCTT GAAATGTAT CTCTCAGTAA TTTTGAAGT CTTTAAAAA ATTTAAAAAC	2340
45	AAAGTGTAG ACTGTGTGG TGTGCGTTGA TGGGCTCA AGAGTCTTCT GATTCATCCA	2400
	GCCCTGCTT TCCCTGGGC CCCCATCCTC TCAGCTCCCG CCGTGGCTTC ACTTGGGGAC	2460
	CCTGCTCGT GTGCTCTTA TGTGCTATT ACTCAGCTA AGGAACACAG TACACTCCAC	2520
50	ACATGCATAA AGGAAATCAA ATGTTATTTT TAAGAAATG GAAATAAAA ACTTTATAAA	2580
	CACCAAAAAA AAAAAAAAAA ACCCGGGGG GGGGCGGTA ACCCATTTCC CTTAA	2635
55		

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 994 base pairs

60

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GAATTCGGCA GAGGTCGGC GAAGATACGG AATAAGGAAG CACAGGAGTA GGGGAGAAGG 60
 AAGCACAGGA GTAGGGGAGA TATACAGCGG TCAGGATAAG GGGGAAAGCG CGGTGGTTGC 120
 10 SCAGAGGTG AAACAAGATG TGAGAGACAA GGGGTAGCGA AGAAATGGGG CAGCGGTTAG 180
 GTTCAGAAGC GCATAGACCG TGCCGGACGG GCAATGCGAG GGGCACAGAA AGGAAGTGA 240
 15 GGGTGGGCTA TTTTAARGGA GATGGTCCTT CAGCCCTCTT YTTTTCTGCG TAGTCTCTCT 300
 CCTCCAGGCC GCGCGGGGAT ATGTGCTCGG GAAACCAGCC CAGTCTAGGC TGGATGATGA 360
 CCCACCTCCT TCTACGCTGC TCAAAGACTA CCAGAATGTC CCTGGAATTG AGAAGGTTGA 420
 20 TGATGTCGTG AAAAGACTCT TGTCTTTGGA AATGGCCAC AAGAAGGAGA TGCTAAAAAT 480
 CAAGCAAGAA CAGTTTATGA AGAAGATTGT TGCAAACCCA GAGGACACCA GATCCCTGGA 540
 25 GGCTCGAATT ATTGCTTGT CTGTCAAGAT CCGCAGTTAT GAAGAACACT TGGAGAAACA 600
 TCGAAAGGAC AAAGCCACAA AACGCTATCT GCTAATGAGC ATTGACCAGA GAAAAAGAT 660
 GCTCAAAAC CTCCGTAACA CCAACTATGA TGTCTTTGAG AAGATATGCT GGGGGTGGG 720
 30 AATTGAGTAC ACCTTCCCCC CTCTGTATTA CCGAAGAGCC CACCGCCGAT TCGTGACCAA 780
 GAAGGCTCTG TGCATTCGGG TTTTCCAGGA GACTCAAAAG CTGAAGAAGC GAAGAAGAGC 840
 35 CTTAAGGCT GCAGCAGCAG CCCAAAAAGA AGCAAAAGCG AGGAACCCAG ACAGCCCTGC 900
 CAAAGCCATA CCAAGACAC TCAAGACAG CCAATAAATT CTGTTCAATC ATTTAAAAAA 960
 40 AAAAAAAAAA AAAAAAAAAA AAAAAGGGGA GGGG 994

45 (2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1542 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

GGCASAGCCA COTCGGCCCC GGGCTCCGAA GCGGCTCGGG GCGGCCCCTT CGGTCAACAT 60
 55 CGTAGTCCAC CCCCTCCCCA TCCCCAGCCC CCGGGGATTC AGGCTCGCCA GCGCCCAGCC 120
 AGGGAGCCCG CCGGGAAGCG CGATGGGGG CCGAGCCGCC TCGCTCTGCT TCCTGCTCCT 180
 60 CCTGTTCCGC TCCTGCTGGG CCGCCGCGGG GGGCAAGCTC TCCCAGGACG ACAGCCAGCC 240

376

	CTGGACATCT GATGAACACG TGGTGGCTGG TGGCACCTGG GTGCTCAAGT GCCAAGTGAA	300
5	AGATCACCAG GACTCATCCC TGCAATGGTC TTAACCTGTC TCAGCAGACT CTCTACTTTG	360
	GGGAGAAGAG AGCCCTTCGA GATAATCGAA TTCAGCTGGT TAMCTCTACG CCCCACGAGC	420
	TCAGCATCAG CATCAGCAAT GTGGCCCTGG CAGACGAGGG CGAGTACACC TGCTCAATCT	480
10	TCACTATCCC TGTCCGAATC GCCAAGTCCC TGCTCACTGT GCTAGGAATT CCACAGAAGC	540
	CCATCATCAC TGGTTATAAA TCTTCATTAC GGGAAAAAGA CACAGCCACC CTAAACTGTC	600
15	AGTCTTCTGG GAGCAAGCCT GCAGCCCGGC TCACCTGGAG AAAGGGTGAC CAAGAAGTCC	660
	ACCGAGAACC AACCCGCATA CAGGAAGATC CCAATGGTAA AACCTTCACT GTCAGCAGCT	720
	CGGTGACATT CCAGTTTACC CGGGAGGATG ATGGGGCGAG CATCGTGTGC TCTGTGAACC	780
20	ATGAATCTCT AAAGGAGCT GACAGATCCA COTCTCAACG CATTGAAGTT TTATACACAC	840
	CAACTGCGAT GATTAGGCCA GACCCTCCCC ATCCTCGTGA GGGCCAGAAG CTGTTGCTAC	900
25	ACTGTGAGGG TCGCGGCAAT CCAGTCCCCC AGCAGTACCT ATGGGAGAAG GAGGGCAGTG	960
	TGCCACCCCT GAAGATGACC CAGGAGAGTG CCTGTATCTT CCTTTCTCTC AACAAGAGTG	1020
	ACAGTGGCAC CTACGGCTGC ACAGCCACCA GCAACATGGG CAGCTACAAG GCCTACTACA	1080
30	CCCTCAATGT TAATGACCCC AGTCCGGTGC CCTCCTCCTC CAGCACCTAC CACGCCATCA	1140
	TGGGTGGGAT CGTGGCTTTC ATTGTCTTCC TGCTGCTCAT CATGCTCATC TTCCTTGGCC	1200
35	ACTACTTGAT CCGGCACAAA GGAACCTACC TGACACATGA GGCAAAAGGC TCCGACGATG	1260
	CTCCAGACGC GGACACGGCC ATCATCAATG CAGAAGGCGG GCAGTCAGGA GGGGACGACA	1320
	AGAAGGAATA TTTTATCTAG AGGCGCCTGC CCCTTCTCTG CGCCCCCAG GCCCCTGTGG	1380
40	GGACTTGCTG GGGCCGTGAC CAACCCGGAC TTGTACAGAG CAACCGCAGG GGGCGSCCT	1440
	CCCGMTGTGT CCCCAGCCCA CCCACCCCTT TGTTACAGAA TGTATKGTTC GGGGTGCGGT	1500
45	TTTGTWATTG GTTTNGGATN GGGGAAGGGA GGGANGGCGG GG	1542

(2) INFORMATION FOR SEQ ID NO: 124:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1390 base pairs

(B) TYPE: nucleic acid

55

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

60

CAAGCTCTAA TACGACTCAC TATAGGGAAA GCTGGTACGC CTGCAGGTAC CGGTCCGGAA

60

377

TTCCCGGGTC GACCCACGCG TCGGGGCTC AGGGTGGACG CATGGTTCTG CACTGAGGCC 120
 CTCGTCATGG TGGCGCCTGT GTGGTACTTG GTAGCGGCGG CTCTGCTAGT CCGCTTTATC 130
 5 CTCTTCCTGA CTGCGAGCCG GGGCGGGCG GCATCAGCCG GCCAACAGCC ACTGCACAT 240
 GAGGAGCTGG CAGGAGCAGG CCGGGTGGCC CAGCCTGGGC CCCTGGAGCC TGAGGAGCCG 300
 10 AGAGCTGGAG GCAGGCCTCG GCGCCGAGG GACCTGGGCA GCGGCTACA GCGCCAGCGT 360
 CGAGCCGAGC GGGTGGCCTG GGCAGAAGCA GATGAGAAGC AGGAGGAAGC TGTCTCTCTA 420
 GCCCAGGAGG AGGAAGGTGT CGAGAAGCCA GCGGAAATC ACCTGTGCGG GAAATTTGA 480
 15 GCTAAGAAAC TCGGAANNT GGAGGAGAAA CAAGCGCGAA AGGCCAGCK TGAGGCAGAG 540
 GAGGCTGAAC GTGARGWGG GAAACGACTC GAGTCCGACC GCGAATGAGT GGAAGAAGCA 600
 GGAGGAGCGG CTTCGCCTGG AGGAGGAGCA GAAGGAGGAG GAGGAGAGGA AGGCCCGCGA 660
 20 GGAGCAGGCC CAGCGGGAGC ATGAGGAGTA CCTGAACTG AAGGAGCCCT TTGTGGTGA 720
 GGAGGAAGGC GTAGGAGAGA CCATGACTGA GGAACAGTCC CAGAGCTTCC TGACAGAGTT 780
 25 CATCAACTAC ATCAAGCAGT CCAAGTTGT GCTCTTGAA GACCTGGCTT CCCAGGTGG 840
 CCTACGCACT CAGGACACCA TAAATCGCAT CCAGGACCTG CTGGCTGAGG GCACTATAAC 900
 AGGTGTGATT GACGACCGGG GCACTTCAT CTACATAACC CCAGAGGAAC TGGCCGCGCT 960
 30 GGCCAACTTC ATCCGACAGC GGGGCGGGT GTCCATCGCC GAGCTTGCCC AAGCCAGCAA 1020
 CTCCTCATC GCCTGGGGCC GCGAGTCCCC TGCCCAAGCC CCAGCCTGAC CCCAGTCTT 1080
 35 CCCTCTTGA CTCAGAGTTG GTGTGGCCTA CCTGGCTATA CATCTTCATC CCTCCCCACC 1140
 ATCCTGGGA AGTGATGGTG TGGCAGGCA GTTATAGATT AAAGCCCTGT GAGTACTGCT 1200
 GAGCTTGGTG TGGCTTGGTG TGGCAGAAGG CCTGGCCTAG GATCCTAGAT AAGCAGGTGA 1260
 40 AATTTAGGCT TCAGAATATA TCCGAGAGGT GGGGAGGGTC CCTTGAAGC TGGTGAAGTC 1320
 CTGTTCTTAT TATGAATCCA TTCATTCAAG AAAATAGCCT GTTGCAAAAA AAAAAAAAAA 1380
 45 AAAAATCGA 1390

50 (2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1288 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

60 GCGCGCGCGG TGAAAGGCGT ATTGATGCGG CTGCGGCGG CCTCGAGCG CCGCGGAGCA 60

378

5 GACGCTGACC ACGTTCCTCT CCTGGGTCTC CTCCGCTCC AGTCCGGCCG TGCCCGCCAG 120
 CCGGGAGCCA TGCGACCCCA GGGCCCCGCC GCCTCCCCGC AGCGGCTCCG CGGCTTCTG 180
 CTGCTCTCTG TGCTGCAGCT GCGCGCGCCG TCGAGCCCT CTGAGATCCC CAAGGGCAAG 240
 CAAAAGGCCG ATCCGGCAGA GGGAGGTGGT GGACCTGTAT AATGGAATGT GCTTACAACG 300
 10 GCCAGCAGGA GTGCTTGGTC GAGACGGGAG CCCTGGGGCC AATGGCATTG CCGGTACACC 360
 TGGGATCCCA GGTCCGGATG GATTCAAAGG AGAAAAGGGG GAATGTCTGA GGGAAAGCTT 420
 TGAGGAGTCC TGGACACCCA ACTACAAGCA GTGTTTCATG AGTTCATTGA ATTATGGCAT 480
 15 AGATCTTGGG AAAATTGGG AGTGTACATT TACAAAGATG CGTTCAAATA GTGCTCTAAG 540
 AGTTTTGTTC AGTGGCTCAC TTCGGCTAAA ATGCAGAAAT GCATCTGTG AGCSTTGGTA 600
 20 TTTCACATTC AATGGAGCTG AATGTTCAAG ACCTCTTCCC ATTGAAGCTA TAATTTATTT 660
 GGACCAAGGA AGCCCTGAAA TGAATTCAAC AATTAATATT CATCGCACTT CTTCTGTGGA 720
 AGGACTTTGT GAAGGAATTG GTCTGGATT AGTGGATGTT GCTATCTGGG TTGGCACTTG 780
 25 TTCAGATTAC CCAAAAGGAG ATGCTTCTAC TGGATGGAAT TCAGTTTCTC GCATCATAT 840
 TGAAGAACTA CCAAAATAAA TGCTTTAATT TTCATTGCT ACCTCTTTTT TTATTATGCC 900
 30 TTGGAATGGT TCACTTAAAT GACATTTTAA ATAAGTTTAT GTATACATCT GAATGAAAAG 960
 CAAAGCTAAA TATGTTTACA GACCAAAGTG TGATTTTACA TGTMTTTAAA TCTAGCATTA 1020
 TTCATTTTGC TTCAATCAA AGTGGTTTCA ATATTTTTTT TAGTTGGTTA GAATACTTTC 1080
 35 TTCATAGTCA CATTCTCTCA ACCTATAATT TGGGAATATT GTTGTGGTCT TTTGTTTTTT 1140
 CTCTTAGTAT AGCATTTTTA AAAAAATATA AAAGCTACCA ATCTTTGTAC AATTTGTAAA 1200
 40 TGTTAAGAA TTTTTTTATA TCTGTTAAAT AAAAATTATT TCCMACAACC TTAACAAAAA 1260
 AAAAAAAAAA AAAAAAAAAA AAAAAA 1288

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1517 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

AGTGGCTTAA AGGCATCGTT TTAGGGATTA CTGGGAAGTA TCTTCAAAGT AATACATGAG 60
 AAACATTCCT TCCTAAATCC TTTATTATAT TGAATATCST ATTAATGGGT TTTCAGAGGT 120

TAAATTAACC ATGTATTCCT GCAATAAATG TCACTTGTTT CTTGTATATA ATCTTTTTTA 180
 TATATTACCG GATTGATTCA TTAGTATTTT GTTGAGGATT TTTGTGTCTA TATTCATAAG 240
 5 AGATGCTGGT GTGCAGTTT CTTTTTTTGT GATAATCTGG TTTTGTATC AGTAATACAG 300
 GCCCCATGAA ACGAGTTGGG AAGTGTTTAC CTCTCTTGTA TTTTTCAG AGTTTGTGAA 360
 GAATTGCTAT TAATTCTTAA AATGTTTGGT AGAATCTACC ATTGAAATCA TGTGTCTCGG 420
 10 GCTTTTTTTT GAGGGAAGTG TTCTGATAAC TAATTCAGTA TCTACTTTTT ATAGCTCTGT 480
 TCAGATTTTG CTCTTCTG AGTTAGTTTT GGTAATTTGT GTATCTCTAG GATTTTGTCC 540
 15 ATTTCAATTA TCTCATTTGT TGCCATAAAT TAACTAAAT TTGGCCTGAG CCTACCTGTA 600
 TATCTTGAGT CCGTCTGTAA GGAAGTGTAG CCTAACTTGT ACATAAACAA ACTGAAATCC 660
 TAAATTAGGA ATGTAGTTTT TGTAACAGCT CCTGAGTCTC AGGCAGTCAC AGCAGYCAAG 720
 20 TCTGTCAATT GCAGGCTGCT AACTAAGCAG CCCATGSTCA AATGAGGCAA AAACCTTTGC 780
 TTTTAACACA TAGTATAGCT TTGTAATCCT TTTCTTGCAC ACTCGGATAA TTTCTTCTTT 840
 25 TTTCAATCCC KGWATTTTCC AKGAATATGA RTCTYCCTTT TTTCCCTCC TGTCACTCTA 900
 GCTAATGGTT TGTCAATTT GTTGATCTTT TGAARAACAA ACCTTTGGTT CCACTTTCTT 960
 GTTGCAATG CTGATATTC TCATTAATGG AGTGGAAGC TGATCTTTGA TTACTTATTT 1020
 30 TACTTAGGGC TGAGGAGTTC ATGGACTTCG CAAAACCTCC TTGAATCTAA ATTGCATCTT 1080
 CTTTCCCTGGT TTCTGGGCTG AAACATGTTT TTTCCCATCT WANAWACCCT TGGTCTTTTC 1140
 35 ATKGGCGATT AAGACTAGAG AAAGTTCTAG ATMCCTTGTG CTTTTATGCT GTCATTTTGT 1200
 TTAAAGGCTT TCTATGTAGT AAACTATCT ATATAGACAA AATAGAGCCT TGACTTGTGG 1260
 TCTTGAATTT GATCAACATG ATTTACCACA TTCTGTACTG GATATTTCTT CACCTGCTGC 1320
 40 TACTGTAAAC CATTTATTC TTGGATCTTC TGTAGAGTAT ATTATCACAG GTACTTTTTA 1380
 CAGGGGTGTC TAATCTTTTG GCTTCCCTGG GCACATGAA AGAAGAAGAA TTGTCTTGGG 1440
 45 CCACACATCA AATACGCTAA CACTAATAAT AGTTGATGAG CTAAAAAAA AAAAAAAG 1500
 GCAAAAAAGN CCCAAA 1517

50

(2) INFORMATION FOR SEQ ID NO: 127:

55

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

380

5 TGPATCTATT CTMTGAACAT TCTACACAA GAATTACATT ATACTGTTAT ACCAGAGTRC 60
 TTCTGCAGTG TGAANTAGAT TGGTTTGGAA AATGAACCTG GCTTTGCTAT AAATTACATT 120
 CACAGGCCTT TTTGCAAAATG TGTAACTTGC CTATCAAAGT AGTTTGTAGG GCAAATGCCAG 180
 AATATATGTC TCCATCTGGT AAAGTACCTT WTAYTCATGT GGGAAATCAA GTAGTATCAG 240
 10 AACTTGCTCC AATAGTCCAA TTTGTTAAAG CCAAGGGCCA TTCTCTTAGT GATGGGCTGG 300
 AGGAAGTCCA AAAAGCAGAA ATGAAAGCTT ACATGGAATT AGTCRACAAT ATGCTGTTGA 360
 CTGCAGAGCT GTATCTTCAG TGGTGTGATG AAGCTACAGT AGGGRMGATC ACTCATGMTA 420
 15 GGTATGGWTC TCCTTACCCT TGGCCTCTGW WTCATATTTT GGCCTATCAA AAACAGTGGG 480
 AAGTCAACG TAAGNTGAAA GCTATTGGAT GGGGAAAGAA GACTCTGGAC CAGGTCTTAG 540
 20 AGGATGTAGA GCAGTGCTGT CAAGCTCTCT CTCAAAGACT GGAACACAA CCGTATTTCT 600
 TCAATAAGCA GCCTACTGAA CTTCAGGCAC TGGTATTTGG CCATCTATAC ACCATTCTTA 660
 CCRACAAAT GACAAATGAT GAACTTTCTG AGAAGGTGAA AAATATAGC AACCTCCTTG 720
 25 CTTTCTGTAG GAGAATTGAA CAGCACTATT TTGAAGATCG TGGTAAAGGC AGGCTGTCTAT 780
 AGAGTTATGT GTTAGTCTCA GGAGCTTAA CTTTTGAAAT ATGTTTTACT TGAATGTTAC 840
 30 ATTAGATATT GGTGTCAGAA TTTTAAACC AAATTACTGC TTTTGAAC CTCAAATTAT 900
 ATAATGTATC TTATGTATGT GCTTTATATT GTTATTGTG TATACATTAA AATAATTCTG 960
 AATTATTTAA TCTGATATGT TGTATTCTGT ATCTTGAAAT TTTTGTFTCC TTGAAACATG 1020
 35 CATGCATTTA AAAATAAAGC TTAACAACCT GTAAAAAAA AAAAAAAA CTC 1073

40

(2) INFORMATION FOR SEQ ID NO: 128:

45

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55

60

CAACCCCTGC CTTTTTTTTG TTTTCCATTT GCTTGGTAGA TCTTCCTCCA TCCCTTTATT 60
 TTGAGCCTAT GTGTGTCTCT GCCCCTGAGA TGAGTCTCCT GAATACAGCA CACTTACTGG 120
 TCTTGACTCT GTATCCAATT TGCCAGTCTG TGTCTTTCAT TTGGAGCAAT TAGCCCATTT 180
 ACATTTAAGG TKAATATTGT TATGTGTGAA TTTRATC/TR TCATTATGWT GTTAGCTGGT 240
 TATTTTGCTT GTTAGTTGAT GCAGTTTCTT CCNGGCATCA ATGGTCTTTA CAANTTGGCA 300

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1275 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

GGCAGAGCCT GTCCCTGCTG CCCCTGCAAA AAAAACCCCC TCTGGTGTGA GCAGGATGGT 60
TGGAGGTTAT GTGAGCTCCT TCTCCTTTCC TCCAGTTTCC TCTTCCCTTC TCCTCCCTGC 120
CTCTTTTGCT TTTCCCTTTC TTCTCTGTAC CCCCTGCCCA TTCTGTATT TTCTCCCATC 180
GCCATTCTCC CCTCTCCAC TGTCCCTAAC CCGTTCAAAC TCTTTCCTCT TAAATGGTTG 240
AGATTTTCTC TCACCAAGCA CACCCAGTA TTAATTAAAC TAGCTGCAA CAGGCAGCAA 300
GTGGTCTACC ATGACAGATG GGTPTTGTGT GTGTGTGTGT GTGTGTAATT GTAATAAAAC 360
ATATTGARTC ACTCAATAAA CACAGAGTGT CTA CTACTACATG TATCARGCAC TATCATAGAT 420
GCTAATTAAC GAAACTGAAA TGGCCAGGCC CTCACAGTGG CTCATGCCTA TAATCCCAGC 480
ACTTTGGGAG GATGAGGCAG GAGGATCACT TGAGGCCGGG AGTTCAAGAC CAGCCTGGGC 540
AACATAGTAA GACTCCATCT CTACAAAAA AAAATTTTTT TTATTATACT TTAAGTTTTG 600
GGTTACATGT GCAGAACGTG TAGTTTTGTT ACATAGGTAT ATACGTGCCC TGGTAGTTTG 660
CTGCACCCAT CAACCCATCA CCTACATTAG GTATTTCTCC TAATGTTACC CCTCTCCTAG 720
CCCCCACCC CGTGACAGGC CCTGCTGTGT GATGTTCCCC TCCCTGTGTC CATGTGTTCT 780
CAATTGGTCAA CTCTCACCTA TGGAGTGAGA ACATGTGGTA TTGGTTTTTC TGATCTTGTG 840
ATAGCTTGCT GAGAATGTG GTTTCCAGCT TTATCCACGT CCCTGCAAAG GGCATAAACT 900
CATCCCTTTT TATGGCTGCA TAGTGTTCCA TGGTGATAC GTGCCACATT TTCTTAATCT 960
ATCATTGATG GACAAGTTTT GCTATTGTGA ATAGTGCCAC AATAAACATA CGTGTGCGTG 1020
TGTCTTTATA GCAGCATGAT TTATAATCCT TTGGGTATAT ACCCAGTAAT GGGATCACTG 1080
AGTCAAAATGG TAFTTCTCGT TCTAGATCCG TAAGGAATTG CCACACTGTC TTCCACAATG 1140
TTTGAACATA TMTACACTCC CACCAACAGT GTAAAAGTGT TTCTATTTTT CCACAACCTC 1200
TCCAACATCT GTTATTTCTT GACTTTTAA TGAACGTGAT TCTAACTGGC GTGAGATGGT 1260
ATCTCATTGT GGTTC 1275

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CNGAAACCCC GTGAACCCCTC CCCGGGTAA AAAGCCCCCC CTAAATGGGG GGAACGCYTC 60
ACACGTTATA AAAAAGCACT AGAATGTTTT GAAAGCGAGA AACAACAGCT GTGTAGGGTA 120
GCTAGCAGTT AGTGTGTAC AGAAGACAGA TATTTGTGCA TTTVTGCATT TTCTAAGTTT 130
GCTGCAATGA GCATGTATTA CTTTCATAGT TATAAAACAC ATGCAAAATG CCCTTTTAAA 240
ATGAAAAAAA ATCCATGAGT GTAAGTGATA TATATGCTTT GGAAAGCCTG GGACGGTCAT 300
TGTTTACTCT CAATAGTATG TGTTTGCCTT TGTCTTTTTC AGACATTTTC TTTTAATCTG 360
TTGATGACAA TAACCTGTTG ATAATATAAC TTGATAACAA ATAAAATGAC TTATGATTGA 420
AWMAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA NN 472

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1950 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

ACCTCTCAGA ATCTTCTCTC AGCAACCTGA GTCTTCGCCG TTCCTCAGAG CGCCTCAGTG 60
ACACCCCTGG ATCCTTCCAG TCACCTTCCC TGGAAATTCT GCTGTCCAGC TGCTCCCTGT 120
GCCGTGCCTG TNAATCGCTG GTGTATGATG AGGAAATCAT GGCTGGCTGG GCACCTGATG 130
ACTCTAACCT CAACACAACC TGCCCTTCTT GCGCCTGCCC CTTTNTGCCC CTGCTCAGTG 240
TCCAGACCTT TGATTCCTGG CCCAGTGTCC CCAGCCCCAA ATCTGCTGGT GCCAGTGGCA 300
GCAAAGATGC TCCTGTCCCT GGTGGTCTTG GCCCTGTGCT CAGTGACCGA AGCTCTGCCT 360
TGCTCTGGAT GAGCCCCAGC TGTGCAACGG GCACATGGGG GGAGCCTCCC GGCGGTTGA 420
GAGTGGGGCA TGGGCATACC TGAGCCCCCT GGTGCTGCGT AAGGAGCTGG AGTCGCTGGT 480
AGAGAACGAG GGCAGTGAGG TGCTGGCGTT GCCTGAACTG CCCTCTGCCC ACCCATCAT 540
CTTCTGGAAC CTTTGTGGT ATTTCACAG GCTACGNCCT CCCAGTATTC TACCAGGCCT 600
GCTGCTGGCC TCCTGTGATG GGCTTCGMA CTCCCAGGCC CCATCTCCTT GGCTAACCCC 660

5
 10
 15
 20
 25
 30
 35
 40
 45

TGATCCAGCC TCTGTTGAGG TACGGCTGCT GTGGGATGTA CTGACCCCTG ACCCCAATAG 720
 CTGCCCACCT CTCTATGTGC TCTGGAGGCT CCACAGCCAG ATCCCCCAGC GGGTGGTATG 780
 GCCAGGCCCT GTACCTGCAT CCGTTAGTTT GGCACGTGTG GAGTCAGTGC TGGCCCATGT 840
 TGGACTCAAT GAACTGCACA AGGCTGTGGG GCTCCTGCTG GAAACTCTAG GGGCCCCCAGC 900
 CACTGGCCCTG CACCTGCAGA GGGGAATCTA CCGTGAGATA TTATTCTGTA CAATGGCTGC 960
 TCTGGGCAAG GACCACGTGG ACATAGTGGC CTTGATAAG AAGTACAAGT CTGCCTTTAA 1020
 CAAGCTGCCC AGCAGCATGG GCAAGGAGGA GCTGAGGCAC CGGCGGGGCG AGATGCCCCAC 1080
 TCCCAAGGCC ATTGACTGCC GAAAATGTTT TGGAGCACCT CCAGAATGCT AGAGACCTTA 1140
 AGCTTCCCTC TCCAGCCTAG GGTGGGAAG TGAGGAAGAA GCGATTCTAG AGTTAAACTG 1200
 CTTCCCTGTT GCCTTCATCG AGTTGGGAAC AGGCTGGGAA GGATGCCCAG TCAAAGGCTC 1260
 CAAGCGAGGA CAACAGGAAG AGGGATCCAC TGTTACCAA AGTCCTGATT CCCCCATCAC 1320
 CAACCTACCC AGTTTGTTCG TGCTGATGTT GGGGGAGATC TGGGGGAGT TGGTACAGCT 1380
 CTGTTCTTCC CTGTCTCTAT ACCGGGAAGT CCCCTCCAGG GTACCCACAG ATCTGCATTG 1440
 CCTGGTCAT TTTAGAAGTT TTTGTTTTAA AAAACAAGTG GAAAGATGCA GAGCTACTGA 1500
 GCCTTTGCCC TGAATGGGAG GTAGGGATGT CATTCCTCAC CAATAATGGT CCTCTTTCCC 1560
 TGACGTGCT GAAGGAGCCC AAGGCTCTCC ATGCCCTTCT ACCTAAGTGT TTGTATTTTA 1620
 TTTTAAATTA TTTATTCTGG AGCCACAGCC CCGTTGCTTA TGAGGTTCTT ATGGAGAGTG 1680
 AGAAAGGGAA GGGAAATAGG GCACCATGGT CCGGTGTTT GTAGTTCCTT CAAAGTCAGG 1740
 CACTGGGAGC TAGAGGAGTC TCAAGCTCCC CTTAGGAAGA ACTGGTGCCC CCTCCAGTCC 1800
 TAATTTTCT TGCCCTGCCCC GCCTTGGGGA ATGCCCTACC CACCCAGGTC CTGACCTGTG 1860
 CAATAAGGAT TGTTCCTGC GAAGTTTGT TGGATGTAAA TATAGTAAA GCTGCTTCTG 1920
 TCTTTTTCAA AAAAAAAAAA AAAAAAACT 1950

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 990 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TGAAGATTT AAAATAGTT TCATATTTCT CTTGAATATG AATATATAAG CTTGAATAAG

60

384

5
 10
 15
 20
 25
 30

CTGAGTCTT TATTATTAGG AATTTTTCCT TATTATTTCT ACCGATGCTT CTTATATTAA 120
 AGCGTAGCTT TTTCGAAAT AATATATGTA CATTAGCTGC CTGCGGATTA ACATTTCCAT 130
 GAATGTATT TTTCGATCTT TTAGCTTTAA ACTTTTGTG TCTTTATATA AGGTATGCTY 240
 CTTTAAAGA TGAATTTCTT AACGCAATA GTTGAAGAC AATCT/CACC TTCTACTTGT 300
 AATTTTAGCT GTTAGTAAAT TTTTGATGCA TATTACGTCT TATTATTTAA CCAACCTATT 360
 TTTTATATC TAGGGGCTT TTGGAAGAC CTTATTTTCT TGTATTAATC AAATATTTT 420
 AAGATTGTAT TTCTCTTAT TATTAGKAA TACGKTACYC YAAATATATA TTGTGGSTAT 480
 TTTCGAAAT GCAATATGCG TCCTTAATTT ATTAGAGGCT AACCTAAAT ATTACTTTTA 540
 CGACTTACTT GAATTTCTG GAATTTAGA ACATTTATTG TTTTATGCAT TTAAATCTA 600
 CTGTATTTT TACTACTCTT AAACATTATT ATTGTTTTAG ACPAGCCAAA ATATATNTTG 660
 TTTATCTT ATCTGCTT TCTTTCTGTA TTTTATGCC ACTATGTATG CTCAATTTCC 720
 TTCTATGTA TGAACCTAAT TCAGTACTTT TGTTTTTTAA TCTGTGCAGG TAGCCTGGCC 780
 ATTAAATTT TATTTTGGT TTGCTGAAAA AATTGTGTTT ATTTCTATAT GCATACTTAT 840
 GCAATAGAA TACTAGTNG ACATATTTT AGTATTTATA AATGTAAAGT CATTWATTNG 900
 GCTTCTATCA TTCTCTGGA GAATCAATT GTGAGCCCAA TAGTTTTTCA TTTTAAATTA 960
 CAGATTTT TCAATTTCT GTTTTAGGA 990

35

(2) INFORMATION FOR SEQ ID NO: 133:

40
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1720 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

50
 55
 60

GTCTGATAG CGACTGTGGT TATCCCCCTA AAGTTTACTT CAGCACTAAC ACTAGTGCTT 60
 CCGGTGGAGT TTGCAATTTT CCAGCTTTAT ACAGGATTTT CCTTTGACTG GAAGAGTCAA 120
 GCAATAGAG ACTCAAGCTT GACATTTATT GTACACATC AAGGGGAATA GCATACTCAT 180
 CAACTGGGA TTATTTTAT CAAACATCG TCTCTTTGA ATAAGAAAAA TACATAGTTG 240
 GTTATTATGG ACTTAAACT GTGTTAAATG GATATTCTGA TAAATATTT GCTGCTCTGT 300
 AGAGTGTGGA AATCTGGA ATATTAGCTT TACTCATCTT CAGCTTTGAG GATGTTCTCT 360
 GTAGCCCGAT GGTTCATAT TAACTAAAAA AGCTGGGTAT TGTAAATCT CATTATATAA 420
 AACTCAGATG AAGAGAAAT TTTCTTTGAT GGTGAGACTG TTGCTTAGT TCAGGAAAT 480

ATTTAATAAT CCTTTGTTAC CTGTGAATGA AGGAACCTTG TAATCTGTAT TTATCGTAAA 540
 ACATGAGCCT TTCCAGAGTC AGCTTAGACA CTGTTGTCCG AATAGCCAT GCTTTGCCCT 600
 5 ATGCCAAGGA GGGCCAGAGG GAGGGCCTAG TCTTCCTCTG TTGCTGTACA TATATTGAAA 660
 TGCTTTTTTT TTTTATTTTG CATTTGTTAT CTATAATGAG CTTTCTGAGC CTTGATATTA 720
 10 TGTGAGACAA ACAGGAGTTA TTGATGTTAT ACACTCCCTT CCATTCAGGA TTTTCTGCTT 780
 GGAGGGAAAT ATGTTGACCT TAGAGAATTG TGAATATTGT TGCAATTCTT GAATATATTA 840
 CCATGTGAAT AATAGAGACT GTGTTGCTCT CTAGTATAAG CTATATTTAT TTTTGATTCA 900
 15 TTTGAATTAC TAGTTATAAC TGGAGAAATT TTGTTACCTC TATCCTGGCT TGCTTGACTG 960
 GCTGTATAAT AGCAGCAGCC TCTTTTAGAG CATCTTAATG AAAACATGGA TGAAAGGAAT 1020
 20 TAATGATGAT ATCTGCAGAC TGCCTAGAAA ATGGCTTTTG TTCCGAGCGT TAACATTTTC 1080
 TTCTCAATCA CATTTCAATG TTTGTGGAGA GTGGCAGATT CACACCAGAA ACACTAGGTG 1140
 TTCATATCCA TAGCATGGAT GCAGAATAAG CAGTTGGGAG AGAAGCTTCT TCCTACCTGG 1200
 25 TACTCCTCCC ATTCACCTCA GCCCAGCCCC AGACAGGCGT TAGCATTCAG TGTGGGGCCT 1260
 CAGGCAGCCC TGAAGCCTGG CTGGGTGATC AGATGGGGGC AGCCTGTGAC GGGCACCAGC 1320
 30 GGCCTGATTC CAGGGAAGAG TTCCTGGAGG GTGTTGGCTG TTTTGTGTAG CTCAGTTTTT 1380
 TTCTGGGCTC CACCATTCTT AACTCCAGGT AGACAAGATA GATGTCACAC ACAACAATTT 1440
 TAAAGTATTT TGCTTAGTGC ATTTTGTGTA TGATTGCAGT GTTTGTTTCT TATTTAATAG 1500
 35 GCTTTTTACT TCATTCTATT AAATTTTAGT GTTTAGAAGA GCGGGGTACT GTCACTGTGT 1560
 AAAATATGTA ATATTTTATA TGTTATACCA TGTCATATAT ACTTGCAATA TCAGACCTTG 1620
 40 CATTCATAT ACAATGCAAT TGACTCTTTG CAGACCTGCA TTTTTCAGTG AACATAAAA 1680
 AGATTGTCTG GCACTCCAAA AAAAAAAAAA AAAAAAAAAA 1720

45

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 705 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

55

GGCAGGAGGC CATCTGGGCT CATTCAGCAG GAAATAATGG AAAAGCTGC AATATCCAGG 60

TGTTTACTAC AATCTGGAGG CAAGATCTTT CCTCAGTATG TGCTGATGTT TGGGTTCCTT 120

60

386

GTGGAATCAC AGACRCTOCT AGAGGAGAAT GCTGTTCAAG GAACGGAACG TACTCTTGGA 180
 TTAAATATAG CACCTTTTAT TAACCACTTT CAGGTACCTA TACGTGTATT TTTGGACCTA 240
 5 TCCTCATTCG CCTGTATACC TTAAAGCAAG CCAGTGGGAA TCCTAAGACT AGATTTAATG 300
 ACTCCGTATT TGAACACCTC TAACAGAGAA GTAAAGGTAT ACGTTTGTNA AATCTGGGAA 360
 GACTTGACTG CTATTCATT TTGGGTATCA TATGTACCTT GATGAAGANG ATTAGGTTGG 420
 10 GATACTTCAA GTGAAGCCTC CCACTGGAAA CAAGCTGCAG TTGTTTTAGA TAATCCCATC 480
 CAGGTTGAAA TGGGAGAGGA ACTTGTAATC AGCATTACG ATCACAAGAG CAATGTCAGC 540
 15 ATCAGAGTAA AGCAATGAAG AGCAGTTTTT CATGAAAAC TGTGTAAATA GAGCATCAAC 600
 AAGTACAAAA TTCTTGTCTT AATTAGTGGG GGTATATAAA AATTCCTTGT AATGGTCAAA 660
 20 TATTTTTTAA AATTGACATT AATAAAGCAT ATTTTAAAG TTTCT 705

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

AGCACACACC TCCTTAGTT GCTCCTAAGG TCATGTTCAA CATTCGTGGA GTGCATTTTC 60
 35 TGCTCAGGGA GCTTCCCAG ACCCGGAATG TTTGGTGCTC ACAGACYCTG GCAAGGATCG 120
 GTATTGCTGT TCCTCAGTTT TGCCTGGGGA AATGGAGGST CAGTGACGTT CAGTGACGTG 180
 40 CCCAGAGTCA TGCCATTGGC GGGTGGCCCA GCGMTCCAGG TCTCCAGCAC CCCTCGGCCC 240
 CCTCCTCACC AGGTACATC ATCTCCTGGA TTAGAATCTG CTCACATAGT CTGTCTGAA 300
 45 AGGAAAAAAA AAAAAAAAAA AAC 323

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 582 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

GGACGGAATG GTGCAACCCT COTWAMTTTT CTRKKSCTGT TCACACAGA GGGAGGGAGG 60

5 GAAAACATTT TTYGTGGGAG AATCCTACTT CTGCAGSGGA GGCCTTAAGC GATKGATTTT 120
 GAATCTKGAC CCTTTACCPA CTAATTTTGA AGGAAGATAC CTTGGAPATA TTTGGCATTG 180
 10 AGTGGGTTAC TGAACAGCA TTAGTGAATT CATCTAGAGA ACTCTTTCAT TTATTCAGGC 240
 AACAACTGTA CAACTTGGAA ACCTTGTTAC AGTCCAGTTG TGATTTTGGG AARGTATCAA 300
 CTCTACACTG CAAAGCAGAC AATATTAGGC AGCAGTGTGT ACTATTTCTC CATTATGTTA 360
 AAGTTTTCAT CTTGAGGTAT CTGAAAGTAC AGAATGCTGA GAGTCATGTT CCTGTCCATC 420
 CTTATGAGGC TTGGAGGCT CAGCTTCCCT CAGTGTGAT TGATGAGCTT CATGGATTAC 480
 15 TCTTGATAT TGGACACCTA TCTGAACCTC CCAGTGTTAA TATAGGAGCA TTTGTAAATC 540
 AAAACCAGAT TAAGGTTTGA CTGGTTTCAT TTGATTTTGA AG 582

20

(2) INFORMATION FOR SEQ ID NO: 137:

25 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1021 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

TTGGGCAGAG CCCTTGCGCG CTCTTGAATA CCTGCKTTCT GTAGGGCTAG TTCTCTTCAA 60
 GATTGCTTA GTGTGATTC ATTTGGGTTT CTTTCTCGC CATGTTTTTC TGTGGAATT 120
 35 ACGGTTGTT TTGGTTCTAT GTACTCTCTA AAATGTTATC GTTTTTCATT TGTCTACTAA 180
 TTTTGGTGA TTTGTTACTA CTGAGTTTCT TAATATCTGA CTGGCCTCCG CCCACGGGCT 240
 40 CTGCAGANCA TAAATACTC AGGCTGATGG TAGTGAGAG ACTCTCCCTC CTTGATCAGC 300
 GCAAACGTTG GTCTGAGGCT TGAGGGATGG AGCAACATTT TCTTGGCTGT GTGAAGCGGG 360
 CTTGGGATTC CGCAGAGGTG GCGCCAGAGC CCCAGCCTCC ACCTATGTTG AGTTCAGAAG 420
 45 ATCGTGGGCC GTGGCCTCTT CCTTTGTATC CAGTACTAGG AGAGTACTCA CTGGACAGCT 480
 GTGATTTGGG ACTGCTTTCC AGCCCTTGCT GGCGGCTGCC CGGAGTCTAC TGGCAAAACG 540
 50 GACTCTCTCC TGGAGTCCAG AGCACCTTGG AACCAAGTAC AGCGAAGCCC ACTGAGTTCA 600
 GTTGGCCCGG GACACAGAAG CAGCAAGARG CACCCGTAGA AKAAGTGGGG CAGGCAGARG 660
 AACCOCAGAC ACTCAGGCTC CAGCAGCTTC CTTGGAGCAG TCCTCTCCAT CTTTGGGACA 720
 55 GACAGCAGGA CACCGAGGTC TGTGACAGCG GGTGCTTTT GGAACGCCGC CATCCTCCTG 780
 CCTCCAGCC GTGGCGCCAC CTCCCGGGTT TCTCAGACTG CTTGGAGTGG ATTCTTCGGG 840
 60 TTGGTTTTC CCGCTTCTCT GTACTCTGGG CGTGCTGTTT ACCGATCTCT GGAGCTAAGC 900

AGCCTTAGAT AGCAGCAGAA GGCTTTTGG ATTCTCCTCC TTGAAAAGAT TCTCAGTTAC 960
 CRAACGTCTC CACCTAGAAA ATAAAAATAC ATTAAGATGT TGANPAAAAA AAANAAAAAA 1020
 A 1021

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1777 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

CGGAAGATGA TGGCTTCAAC AGATCCATTC ATGAAGTGAT ACTAAAAAAT ATTACTTGGT 60
 ATTCAGAACG AGTTTAACT GAAATCTCCT TGGGGAGTCT CCTGATCCTG GTGGTAATAA 120
 GAACCATTC AATACACATG ACTAGGACAC GAGACAAGTA CCTTCACACA AATTGTTTGG 180
 CAGCTTTAGC AAATATGTCG GCACAGTTTC GTTCTCTCCA TCAGTATGCT GCCCAGAGGA 240
 TCATCAGTTT ATTTTCTTTG CTGTCTAAAA AACACAACAA AGTTCTGGAA CAAGCCACAC 300
 AGTCCTTGAG AGGTTGGCTG AGTTCTAATG ATGTTCTCTT ACCAGATTAT GCACAAGACC 360
 TAAATGTCAT TGAAGAAGTG ATTCGAATGA TGTTAGAGAT CATCAACTCC TGCCTGACAA 420
 ATTCCCTTCA CCACAACCCA AACTTGGTAT ACGCCCTGCT TTACAAACCC GATCTCTTTG 480
 AACAAATTCG AACTCATCCT TCATTTGAGG ATATAATGCA AAATATTGAT CTGGTGATCT 540
 CCTTCTTTAG CTCAAGGTTG CTGCAAGCTG GGAGCTGACC TGTCAGTGGA ACGGGTCTTG 600
 GAAATCATT AAGCAAGGCT CGTTGCGCTG CCCAAAGACA GACTGAAGAA ATTTCCAGAA 660
 TTGAAATTCA AATATGTGGA AGAGGAGCAG CCGGAGGAGT TTTTATCCC CTATGTCTGG 720
 TCTCTTGTCT ACAACTCAGC AGTCGGCCTG TACTGGAATC CACAGGACAT CCAGCTGTTT 780
 ACCATGGATT CCGACTGAGG GCAGGATGCT CTCCCACCCG GACCCCTCCA GCCAAGCAGC 840
 CCTTCAAGTT CTTTATTTTC TGGTAACAG AAGTAGACAG ACAGGTTACT TGGTGTATCT 900
 TCTGTAAAG AGGATGTCAC GAGTGTGTTT TCCTCAGACA CTTTGATTTG GAGAATTGGT 960
 GCTAGTTGGC AATAGATAAC TCAGCGTAGA TAGTATTGCA AAAAGGGGAG GAAATACACA 1020
 ACAATAATAA ATGTAATAAC CTGCTATTCA ACATGCAGTT TTATTTTGGT GCCAAAAATC 1080
 TAGAGCTTTC CCAAGATCCT GTTGCCCTAG GCACATNCAC ACTTCAACAG TGCACACTAT 1140
 CCAACAGTGC AACTATTCA AAGTGCACA CTATTCAAAA CGGTAGACTA TTTTTTTGCA 1200

389

TGTTCAAGAT ATTTGTTTGG GTCTTATGTG TGTGTGAGAG AGAGAGATTC CTTTGACATT 1260
AAGGAGCATC AATGAGAAA GATGATGAGG CAGGAATTAA TAAAGAAATG AAGTCGTGTG 1320
5 TGTTTGTTTG CCGTGTGAGAG GGCACACAAT TTCATTAACA CCATGCCCTGG ACAATTTGAT 1380
ATTAATATTT AACACCTCTG CATCTTTTTC TTAATAAAGA ATATGGGCCA GATACAGTGG 1440
CTCAGATTTG TAATCCCAGC ACTTTGGGGA GCCAAGTTAG CAGAATCCCT TGAGCACAGG 1500
10 AATCTGAAAC CAGCTTGGGC AACATAGTGA GATCCCATCT NTACAAAAA CTAAAAAAT 1560
AGCCAGGCAT GATGGCACAT TCGTGTAGTC CTAGCTACTC AGGAGGCTAA GGTAGGAGGA 1620
15 TTGCCCTGAGC CCAGGAGTTC AAGGCTGCAG TGAGCTAAGN ACGTGCCAGT AACTCCAGC 1680
CTGAGCCACA AAGTGAGACC CTGTCTCGCA AAAAAAAN TTAATAAGTC GGGGGGGGCG 1740
CCGGTACCCA AATCGCCCGA TATGATCGTA AACATC 1777
20

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 643 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTTTGGG AATGAGAAA TAACTTTATT 60
35 TTCATTTGTG GGAGCGGGCC GATGTCCAGC CTCAGAACTT CTGGAAGTGC TTCTTGGTGC 120
CGGCAGCCTT GGTGACCTTG AGCACGTTGA AGCGCACTGT CTTGCTCAGA GGCCGGCACT 180
40 CCGCCACTGT GACGATGTCA CCGATCTGGA CGTCCCTGAA GCAGGGGGAC AGGTGTACAG 240
ACATGTTCTT GTGGCGCTTC TCGAAGCGGT TGTACTTGGG GATGTAGTGC AGATAGTCTC 300
GGCGGATGAC AATGGTCTTC TGCATCTTCA TCTTGGGTCA CCACGCCAGA GAGGATCCGC 360
45 CCTCGAATGG ACACATTACC AGTGAAGGGG CATTTCTTGT CAATGTAGGT GCGCCTCAAT 420
AGCCTCCTTG GGTGTCTTT GAAGCCCAGA CCGATGTTCT TGTAGTAAC CCGCGGGAGC 480
50 TTCTCCTTGC CAGTTTCTCC CAGCAGGACC CTCTTCTTGT TTTGAAAGAT GGTCCGCTGC 540
TTTTGGTAGG CAGCCTCAGT CTGAATGTCC GCCATCTTCT CGTGCCGMAV TCCTGCAGCC 600
CGGGGGATCC ACTAGTTCTA GAGCGGCCGC ACCCGCGTGG AGC 643
55

(2) INFORMATION FOR SEQ ID NO: 140:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1220 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

10 GGCACGACGA TGATAGACCT ACTGGAGGAA TACATGGTTT ACAGGAAGCA TACCTACATR 60
 AGGCTTGATG GCTCATCCAA GATCTCGGAG AGGCGAGACA TGTTTGCTGA TTTTCAGAAC 120
 AGGAATGACA TCTTTGTGTT CCTGTTAAGC ACACGAGCTG GAGGACTGGG TATCAATCTC 180
 15 ACTGCTGAG ACACAGTGCA TTTTGTATGA TAGCGACTGG AACCCCACTG TGGACCAGCA 240
 GGCATGGAC AGGGCCACC GCTTAGGGCA GACAAAGCAG GTTACTGTGT ACCGGCTCAT 300
 CTGTAAAGGC ACCATTGAAG AACGATTCT GCAAGAGCC AAGGAGAAGA GTGAGATTCA 360
 20 GCGGATGGTG ATTTAGGTG GGAACCTCAA ACCAGATACC TTGAAACCCA AAGAGGTGGT 420
 TAGTCTTCTT CTAGACGACG AAGAGTTGGA GAAGAAACGT ATGTACTCTA AACCTCTATA 480
 25 CACTCCCCTC ACGTATCTGA GAATGGAAGA GCTACTTGGG TGTGTGCCAA GGGTTAGGCA 540
 AAGCCAGAGG CTGTATTTAG GGAAAGTATT TTTGTGCTCA TATTTTATAT AAAAACCCTA 600
 30 ACAAGAATGT GTTTGTAGGC CAGCGTGGT GCGTCGCGCC TCTAGTCTCA GCATTTCCGG 660
 ARGCCAAAGT GGGCAGATCA CCTGARGTCA GGARTTTGAG TTGAPACCA GCGTGGCCCA 720
 CGTTGTGAAA CCCACCTCT ACTARGARTA CSGAAATPG GTTGGGCATG GTGGCGGGCA 780
 35 CCTGTAATTC CAGCACTTTG GGAGGCTGGG GCAGAANAAT TGCTTGAGCC CAGGAGGTGG 840
 AGATTGCGGT GAGCCGAGAT YGTGCCATTG CMTCCAGCC SGGGCAATAA GAGTGAAAYT 900
 CCATCTTTTA AAAACAAACA AAAACAAAAA ACACAAGACG GCTCACACCT GTAATCCCAG 960
 40 CACTTTGGGA RCGCGARGCA GGTGGATCAC GARGTCAGGA GTTCCAAGAC TAGCCTGGCC 1020
 AACCTGGTGA AGCCCCGTCT CTACTAAAAA TACMAATATT AGTCGGCGGT GGTGGTGGGC 1080
 45 ACGTGTATC CCAGCTACTC GGGAGGCTGA GGCAGGAGAA TCCCTTGAAG CTAGGAGGCA 1140
 GAGGTTCAG TGAGCCAGGA TCGTGCCATT GCACTCCAGC CTGGACAACA AGAGCAAGAT 1200
 50 TCCATCTCAA AAAAAAAAAA 1220

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 721 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

5 AATTGGGCAC GAGCCAGGTT AGCCGGAAGG GCAGCTCTCC AGGCCCTGCC CACCCACACG 60
 GGGGCTCCTT ATGCACAGCG GGGGCTCTCC TTGTGGCCAT AGAAACGGAA CTGGCTCTTT 120
 TCAACAGTGC TCCAAGAGGA TGTTATTTTA ACGCTGGCCC CCAAGGAGGA AAGCCACAGA 180
 10 CATTCTCTCC TCCTGGAACA TCCAAGGGCA CTGGATCCTC TGTGTCCCTC TGACATGGGG 240
 TGGCACTCCA GCAAGAGCAC CACGCTGGCA GCTGAGTCCC AGAAGCTTGA AGAAGAGYGC 300
 GAGGGAAGAG AGCCAGGTCT GGAGACCGGC ACCCAGGCAG CAGACTGCCA GGATGCCCCG 360
 15 CTGAAGGATG GAACCCCTGA GCCAAAGAGC TGAATGCCT CTCTCCAGAG TCGGACCTC 420
 ACCTCTTTCC TGGAACTGCC TTTGGCCCCA GAACCATGAG ACAATCCCCA CCTGAGAAG 480
 20 CTCCGATCAC TGGGAGGAGA GAGAAAGCCT CCAGCTTTGG GATTGAGGCT TCAGAAGTTT 540
 TTAGCAGCCT TTGCTCATTC GAGAGGTGGG GAAAGGATAA AGTTCTTATA AGGAAATCCC 600
 TAATTTCCCC CAGCTCTCTC CCNCCNGAAG AAGGAACNAA AGAAAGTTCC TTCCACACGT 660
 25 TTTGTTGGAA ACTTTTCCCT TGCCAACTTT CTTTGGATTG CCAGAACAAA GCCCTCCAGA 720
 A 721

(2) INFORMATION FOR SEQ ID NO: 142:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1468 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

ATGAATTAAT GTTTATAAAT GACTGTACTG AATTTAAAAC CSTACAGTTT CATTTGCATT 60
 45 TTGACATTAC TTTATTATAC ATTTTGCAAT TAAAAGGCTG CACCAGTTGG CTTTTCTTCT 120
 GTTTTATTCT CAAAATATAG AGATTCTGTG ATTTATTTCG COTGTTTATG GATTAAAAAG 180
 AAAATTCTAA TATAAAGCCT TTCAATAGGA TGCATAGGTA TATTACGTTT TTAAATGCT 240
 50 TTAGATCTGT GATTCTTGAC TTAATTATTA TTTTATCCCC TTAAAGTCAG GGATGCTTTA 300
 TTCTATTTTA AAGCACTTAT GAGTTACATG TTGTAATCAA GTTTGCACAA TATATTTATC 360
 55 TATATGAGGA ACCCATAAAT GAATAGCTAA TTTTAAAAAT GCCATTAAAA TGCATGAAAT 420
 KCTTATTAAT ACCTTACTAT ACTATTTCTT CAGGCAAGT AAATTGACCA TGRGAAAGR 480
 60 ACACAGTTAT TAAACACTGT TGACAGGAAA ATTCTCCTTG ATAACATAGG ACAATTAATG 540

392

GAAAAAAAAA TTCTCATTAT TTGCAAAGAA TGAACAAGTT AATGAACAAA CAACTAGAT 600
 TTGGTATGTT TTCAGCTTTT GTATCATGTT TAATTGTTTA ATTGGTTGA AAAACTGCAG 660
 5 TTGAGAAATC AGATAGCAAT ATAGACATTC ACAGCAGCTC TGTGGATACC ATGTAATTGT 720
 CAGGTAATTT CAGAATGTTG AAAATTATTC AGTGCAGCCC TCATAGTATC ATACTTGAAG 780
 10 AAATTGATTA CAGTTCCACT AAATTGTTGA AGATAAATTA TTTTAAAGG TTATGAAAAC 840
 TAAGTTATAT TAATTCATAT GTTGTATTTT TAAATCCAC CTCTCAAGC TATCCAATTT 900
 NCTGACTTTG AAAATAACCA TGAGAGATGC CACATTTCTC TCTGGCAAAC TACCCTCAA 960
 15 AGAATAATTG TTAAAAATTA AGCTTTTAGG TATTGAAGC TGTATAAAG TATAAAATTA 1020
 AGATATAAGC AGATCACATG TAAATCATTC CTAAGCACA AGAAAAGAAT GTGCCTTGAT 1080
 GTACATATAT TACTAAGTTG CCTCTCCCAG TTTACTTTAA AAATGGCTTT AAGGATAAAG 1140
 20 AATAAATGTG ATAGCTGTGC ATGCATTATA TATTTGCATT TGCAAATTTT CCATTGTTTT 1200
 AACAGCTGTG TGGCTGACTT TCAATTTTAA GACGTGAATT GACATACAGC CCATAACTTT 1260
 25 ATAATGGCTG CTCATTTATC TTATCTTTCA GTTAGTGGAA AAACATTTCA ACCTGACTAA 1320
 AATTTGGAAT TGTGTCTTTT ATGTTCCATC CTCTGTGTT ACTAGATTTA GTTTAAAAAT 1380
 TGTGTATGAC CATTAATGTA TGTCAAAAC ATGTAATAA AAGATGTTGA ATCTTGTTGA 1440
 30 AAAGCAWRAA AAAAAAAAAA AAATCTGA 1463

35

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 300 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

45

50

55

TGAATTTTTT GCCAACTTA GTAACCTGT TAAATATTTG GAGGATTTAA AGAACATCCC 60
 AGTTTGAATT CATTTCAAAC TTTTAAATTT TTTTGTACT ATGTTGGTTT TTATTTTCCT 120
 TCTGTTAATC TTTTGTATTC RCTTATGCTC TCGTACATTC AGTACTTTTA TTCCAAACT 180
 AGTGGGTTTT CTCTACTGGA AATTTTCAAT AAACCTGTCA TTATGCTTA CTTTGATTAA 240
 AAAAAAAAAA AAAAAAAAAA AAACCCCNAG GGGGGGCCCC GGTNCCCAAT CCCCCCAA 300

60

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2243 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

5	TGCCTCCCTT CCTGCAGATT GTGGACAGTA GTTCCTCAGC CTGCACCCCTG GATTCCTTCT	60
10	TCCCCTTCTT AGCTCCATGG GACTCGCCCC AAGACTGTGG CTTCAAGGAC CACCAGCCCC	120
	TTACTCTTCA AGCCCTGACT GTGGAGTTGG TAGATGCCTC TGATCCTCAG TATTCTCTCT	180
15	GGCAATGTTT CACGGCTTCT CCTTCCTGGG AGCTGGCTCC ATAACTTGAT TTTCGCCAAA	240
	CGTGTGCAA TCCCTGCTGC CCCTTAGCCA CCCAGGGTCT TGTGTGGGTA TGAGTGTAGA	300
20	GGATGGGGGT ATGCCAGGCC TGGGCCCTCC CAGGCAGGCC CGCTGGACCC TGATGCTACT	360
	CCTATCCACT GCCATGTAGG GTGCCCATGC CCCATTGCTG GCACTGTGCC ATGTGGACGG	420
	CCGAGTGCCC TTYCGGCCCT CCTCAGCCGT GCTGCTGACT GAGCTGACCA AGCTACTGTT	480
25	ATGCGCCTTC TCCCTTCTGG TAGGCTGGCA AGCATGGCCC CAGGGGCCCC CACCCTGGCG	540
	CCAGGTGCT CCCTTCGCAC TATCAGCCCT GCTCTATGGC GCTAACAACA ACCTGCTGAT	600
30	CTATCTTCAG CGTTACATGG ACCCCAGCAC CTACCAGGTG CTGAGTAATC TCAAGATTGG	660
	AAGCAGAGCT GTGCTCTACT GCCTCTGCCT CCGGCACCCG CTCTCTGTGC GTCAGGGGTT	720
	AGCGCTGCTG CTGCTGATGG CTGCGGGAGC CTGCTATGCA GCAGGGGGCC TTCAAGTTCC	780
35	CGGGAACACC CTTCCCAATC CCCCTCCAGC AGCTGCTGCC AGCCCCATGC CCCTGCATAT	840
	CACTCCGCTA GGCTGCTGC TCCTCATTTCT GTACTGCCTC ATCTCAGGCT TGTCGTCACT	900
40	GTACACAGAG CTGCTCATGA AGCGACAGNG GCTGCCCTTG GCACTTCAGA ACCTCTTCCT	960
	CTACACTTTT GGTGTGCTTC TGAATCTAGG TCTGCATGCT GGCGGGGGCT CTGGCCGAGG	1020
	SCTCCTGGAA GGTTCCTCAG GATGGGCAGC ACTCGTGGTG CTGAGCCAGG CACTAAATGG	1080
45	ACTGCTCATG TCTGCTGTCA TGAAGCATGG CAGCAGCATC ACACGGCTCT TGTGGTGTTC	1140
	CTGCTCGCTG GTGGTCAACG CCGTCTCTC AGCAGTCCCTG CTACGGCTGC AGCTCACAGC	1200
50	CGCCTTCTTC CTGGCCACAT TGCTCATTTG CCTGGCCATG CGCCTGTACT ATGGCAGCCG	1260
	CTAGTCCCTG ACAACTTTCA CCCTGATTCC GGACCCTGTA GATTGGGGGC CACCACCAGA	1320
	TCCCCCTCCC AGGCCTTCTT GCCTCTCCCA TCAGCAGCCC TGTAAACAAGT GCCTTGTGAG	1380
55	AAAAGCTGGA GAAGTGAGGG CAGCCAGGTT ATTCTCTGGA GCTTGGTGGT TGAAGGGGTA	1440
	CCCTAGGAG ATGTGAAGTG TGGGTTTGGT TAAGGAAATG CTTACCATCC CCCACCCCA	1500
60	ACCAAGTTCT TCCAGACTAA AGAATTAAGG TAACATCAAT ACCTAGGCTT GAGAAATAAC	1560

5
 10
 15
 20
 25

CCCATCCTTG TTGGGCAGCT CCCTGCTTTG TCCTGCATGA ACAGACTTGA TGAAAGTGGG 1620
 GTGTGGGCAA CAAGTGGCTT TCCTTGCCCTA CTTTAGTCAC CCAGCAGAGC CACTGGAGCT 1630
 GGCTAGTCCA GCCCAGCCAT GGTGCATGAC TCTTCCATAA GGGATCCTCA CCCTTCCACT 1740
 TTCATGCAAG AAGGCCCACT TGCCACAGAT TATACRACCA TTACCCAAAC CACTCTGACA 1800
 GTCTCCTCCA GTTCCAGCAA TGCCTAGAGA CATGCTCCCT GCGCTCTCCA CAGTGTCTGT 1860
 CCCACACCT AGCCTTTGTT CTGGAAACCC CAGAGAGGGC TGGGCTTGAC TCATCTCAGG 1920
 GAATGTAGCC CCTGGGCCCT GGCTTAAGCC GACACTCCTG ACCTCTCTGT TCACCCCTGAG 1980
 GGCTGTCTTG AAGCCCGCTA CCCACTCTGA GGCTCCTAGG AGGTACCATG CTTCCCACTC 2040
 TGGGGCCTGC CCCTGCCTAG CAGTCTCCCA GCTCCCAACA GCCTGGGGAA GCTCTGCACA 2100
 GAGTGACCTG AGACCAGGTA CAGGAAACCT GTAGCTCAAT CAGTGTCTCT WTAAGTGCAT 2160
 AAGCAATAAG ATCTTAATAA AGTCTTCTAG GCTGTAGGCT GCTTCTTACA ACCACAGCCA 2220
 AAAAAAAAAA AAAAAAACTC GAG 2243

(2) INFORMATION FOR SEQ ID NO: 145:

30
 35

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1082 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

40
 45
 50
 55
 60

GCCAAGCTCT AATACGACTC ACTATAGGGA AAGCTGGTAC GCCTGCAGKT ACCGGTTCCG 60
 GGAATTCCCG GGTCGACCCA CCGTCCGCT TCCTGTGTGC AAAATCCTCA CCTCCTTCAT 120
 AACCATCTCC CACAATTAAT TCTTGACTAT ATAAATTTAT GGTTCGATAA TATTATCAAT 180
 TTGTAATCAA TTGAGATTTT TTTAGTGTCT GCTTTTCTGT GACTCAACTG CCCAGACACC 240
 TCATTGTACT TGAAAAGTGG AACANCTTGG GAATGCCATG GGGTTTGATA ATCTGCCAGG 300
 GACATGAAGA GGCTCAGCTT CCTGGGACCA TGAATTTGGC TCAGCTGATC CTGNACATGG 360
 GAGAACAACC ACATTTTCTT TTGTGTGTGC TTCTAGCAGC TGTTCGGGAG GACCKTGACC 420
 CAAYAGTGTT CCCATGCTGT TTCTGTGAA ATGCTCTCGG CTATGTAGCA GCTTTTGATT 480
 CCCTGCATAC CCTAGGCTGC TGCCCCATC CTGTCCCTTG TTTATAACAT TGAGAGGTTT 540
 TCTAGGGCAC ATACTGAGTG AGAGCAGTGT TGAGAAGTCG GGGAAATGG TGAAGTACTT 600
 TAGAGCAAGG CTGGGCATCA GCACCTGTCC AGCTCTACTT GTGTGATGTT TCAGGAACTC 660
 AGCCCTTTT TCTGCCTAGG ATAAGGAGCT GAAAGATTAA CTGGATCTT CTAATGGTCC 720

5 AARTCTTTTG GTCACAATAA AGAGTCTCCA AATTAGAGAC TGCATGTTAG TTCTGGATGG 780
ATTTGGTGGC CTGACATGAT ACCCTGCCAG CTGTGAGGGG ACCCCGTTTT TAAGATGCAT 840
GGCCAAGCTC TCTGCAAAATG GAAATGCTTA CACTGGGTGT TGGGGATGTT TGCTACCTCC 900
TGCTATTTTT GTGGTTTTGG TTCTCCCACT ATGGTAGGAC CCTTGGCCAG CATGTGGGCT 960
10 TGTCTGTCA GCGCCATGTA CTACCTCTC ATGCTCTGAG GTACTACTGC CTCTGCAGCA 1020
CAATTTCTA TTTCTGTCAA TAAAGGAGA TGAAATPAA AANAAAAA AAAAARCTCC 1080
NG 1082

20 (2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 4313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

30 CAAGCTGGTT TGAAACTAGG GGTCCGGCTC GGCCGTCTC GTTGTCTGTC GCGCCATCCC 60
CGCTTCGGGG TTAGGCGGTT COTGCCGGCC CCTCCTCTC CTCCCTCGG ACCCATAGAT 120
CTCAGGCTCG GCTCCCCGCC GCGCCAGCC CACTGTTGAC CCGCCCCGTA CTGCGCCCCC 180
35 CTGGCCACCA TGTCCCTGCA CGGCAACCG AAGGAGATCT ACAAGTATGA AGCGCCCTGG 240
ACAGTCTACG CGATGAACTG GAGTGTGCGG CCGGATAAGC GCTTTGGCTT GCGCCTGGCC 300
AGCTTCGTGG AGGAGTACAA CAACAAGGTT CAGCTTGTG GTTTAGATGA GGAGATTCA 360
40 GAGTTTATTT GCAGAACAC CTTTGACCAC CCATACCCCA CCACAAAGCT CATGTGATC 420
CCTGACACAA AAGGCGTCTA TCCAGACCTA CTGGCAACAA GCGGTGACTA TCTCCGTGTG 480
45 TGGAGGCTTG GTGAACAGA GACCAGGCTG GAGTGTCTG TAAACAATAA TAAGAACTCT 540
GATTTCTGTG CTCCCTGAC CTCTTTGAC TGAATGAGG TGGATCCTTA TCTTTTAGGT 600
ACCTCAAGCA TTGATACGAC ATGCACCATC TGGGGGCTGG AGACAGGGCA GGTGTTAGGG 660
50 CGAGTGAATC TCGTGTCTGG CCACGTGAAG ACCCAGCTGA TCGCCCATGA CAAGAGGTC 720
TATGATATTG CATTTAGCGG GCGCGGGGT GGCAGGGACA TGTTCGCTC TGTGGGTGCT 780
55 GATGCTCGG TCGGGATGTT TGACCTCCGC CATCTAGAAC ACAGCACCAT CATTTACGAA 840
GACCCACAGC ATCACCCTCT GCTTCGCTC TGCTGGAACA AGCAGGACCC TAACCTCTG 900
GCCACCATGG CCATGGATGG AATGGAGGTG GTGATCTAG ATGTCCGGGT TCCTGCACAC 960
60

	CTGTSGCCAG GTTAAACAAC CATCGAGCAT GTGTCAATGG CATTCCTTGG GCCCCACATT	1020
	CATCCTGCCA CATCTGCACT GCAGCGGATG ACCACCAGGC TCTCATCTGG GACATCCAGC	1080
5	AAATGCCCCG AGCCATTGAG GACCCATTC TGGCCTACAC AGCTGNAAGG WGAGATCAAC	1140
	AATGTGCAGT GGGCATCAAC TCAGCCCCAA YGTGCCCCAT CTGCTACAAC AACTGCCTGG	1200
10	AGATACTCAG AGTGTAGTGT TGGTGGCGCT GTGCCCCACGA GGCAGGGGCT TTTGTATTTT	1260
	CTGCCCTCTGC CCCACCCCCA AAGTAAGAAG AAACATGTTT CCAGTGGCCA GTATGTCTTT	1320
	CATTGCTTTG CACCCACTGT TACCAGAAGC TGCTCTAGGA GTTCCTGGCC AGTCACCCCA	1380
15	TGGCCCTCTG TGGCAGACTC AGTGCTGTGT GCGCCCTCCT CAGCCCAGGG CTGAGTTTAA	1440
	AGATTTTCTC TCCTTTCTCT TTCTCCTTTG GTTCCTCAAT TAAAAAATGT GTGTATATTT	1500
20	GTTCCTCAGG CGTTGTGTGT AGGAGCAGTT CACGCCACTG CTGTGTCTAT TCCTCTGCCC	1560
	AGGTGTCTCT GTTTGCTGCC CAAGG/WKKT TTTCACTGTCT CGTCCATGTC CATGTTCTGT	1620
	TTAGCACTWA CGTGGGAACA AATACCAATT TGTCTTTTCT CCTAGTATCA GTGTGTTTAA	1680
25	CAAATTTTAA CTTTGTATAT TTGTATCTA TCAGGCTAAT TTTTATGA AAAGAATTTT	1740
	ACTCTCCTGC TTCATTTCTT TGTCTTATAG TCCTGCCCTCT TTGCACCTTC TTCTCTTCCC	1800
30	TCAGTCCCTG GAGCTGGTAC TGGGCCCCCTG GCCCCATGAG CAGTTTGCCT TCTTGAGTCA	1860
	CTGCCCTGTGT AGTACATACC TGACCCGGAG TCCAAACCCAC CTTGGTGCTC TGAAGTCCAC	1920
	TGACTCATCA CACCTTTCTT AGCCTGGCTC CTCTCAAGGG CATTCTGGGC TTGTAAACAG	1980
35	ACATAGGAAG CCTCTGTTTA CCCTGAAGCA CCACTGTCCA GCCCATTGGT TCCCCTGGC	2040
	AGCATGGTAG AGCTGAGAGA AACAGGCTCT CAGGGTACCT GACTTGAGGG GAATCGTTTC	2100
40	ATGAAGCTGA ACTTCAAGCA TATTTCCAGT ACAITCTTTC AGAGTCTGTT TTTCCATCCA	2160
	AATATAAGCC CCAGGCCATT CCACTTAGTG TCTTTTCAAT GATAGGCAAG AATGATATCT	2220
	GAGTTGAAC TCGGTGCTTC TGTGTTTTGA GTTTACTGTG CCTGGTGGTA TATTGGGCAT	2280
45	TCTTTGGATT GAGTGTCTG AGGTGAGAGA GTCTTCCCGA GGCATCCTGT CTGTGCTTCC	2340
	AACCCTGAAC AAGACCTTAC ATGAGAGATG GACTGATGGA CTGCGGCAAT CCTCGGCTGT	2400
50	CAAGTGGATA GATAGTTAAA AAGCATTATA CTGTGGGTAA TGAAAAGGGA GGAAAAAAA	2460
	AGAAGGAAAA GGAATTATAG ACCCCCAGGG TCAGCCAGTT AAGAGCTCTA CCCACACCTG	2520
	TCAACCCCTC TCTCCCCCAG TTTAGGTTCT GAGCAGTATT GGACTTGTAG CCTGCAGTTG	2580
55	TCTTTGACT TGCAGGCCGC AGTGTCTTTC TGTTATGTGA ATGAGTTCCA TGGAGGGCCA	2640
	TATGTGTGAT TCCACCGTTA GATGAGCCCT TGGGGCAGGC AGTTTGGGAT GTGCTCTTGG	2700
60	GGGAAAGTTG GCTGTTTCTT TGGCTCTGC TCCTACCCGA AGTTTTTAAG TCCCTCTGAA	2760

	TTGCTCATCT GAGATTAGTA GAGTAGCAGG CCTGAAGGAT GATGGTTTTG TCCTCTTTGG	2820
	TTCTGACCTG CTTGAGAAGT AAAACAGTAA CTTTGTTCCT CTGGGCCCTT AAGCTTTTCT	2880
5	GGTTAAGTCT TCCTTTTCAG AAGTAGATGT CATTATATGC CAAAAGTCTA GCTCTTTGCT	2940
	TTACCATACA GGGACCTGTC CCAAGAAAA AGGCTCTTTT TTTAGCCAGC ATATTTCCCC	3000
10	TTCTACCCCTT TTACTTTGTT GTTCTGATTT TAGGACTCTG GCTGGCCCTG TGCTTGTTGGT	3060
	TGCCCTCTCT GCATTTGCCA CTGGATTTGC ACTGCATCGT TTGGAGATAC AAAGCGAGCA	3120
	GTTCTTGGTC AGAACCTTCC TCTGCTTTTC ATTGTGTTTG ATAATGGTTA CTGGGTCTCT	3180
15	CTCTCAAGGG TAGCAAGGCC AAGCTGATGG CTGCTTGTCT AGGAGGCCAT CAGTTCTCTC	3240
	CTGTGGAGAA GGGTCTGAAA TGGAACTCAG TGGTAGAAGG GGCTGGTCTG CTGGGCAGGG	3300
20	CTTACATCCA CTGAGTTCTA AGATTCTTTT CCTGATCTGC ACCTACGCCCT GGTCTGTATG	3360
	GTGGAAATTG TCAGCTGGAA CTCAGAAACA ACAACTTGAA AAAAAATAA TAATTAGAAC	3420
	ATATTTGCAT AAGATAGCTA TTTACTCTGG AAACCAACAA CTTTGTGAGT TTCCCTTGCC	3480
25	CTGTGGACGC CCAGCTCCTG TCATCCTTCC TTAGGTCCTG CAGTACAGTC TTCCCTGAA	3540
	TGCCACCGGG GACCCAGGGG GACTCCACCC CCCTAAGCAA GCACACACAT ACTCACAGTT	3600
30	GATGAGTTGC TGGTCTTTGA GTCCCAGCTC TCTTACCCTC CTTTACTCC ACCAGCCCGA	3660
	CGACCCATGA CTGAGGAGGG GATTTCTAGA GTCTCAGGAT TTAGAAAGTC TGTAAGCCAT	3720
	CCATGTCCA GAAAGCACCG ATCTGTTGTA GTTGCAAAAA CAACTCTGTA ATTTGTTGAG	3780
35	GTTCTCAAAC TGACAGCCAG CGAGACTGGG TGGGAGCCCC TGGATCTGTT CTCCCTGACT	3840
	GCGGGAGGAG CAGCCACTAG GACTTTAGCA GGAAGCCAC ATGGAGGCTC CGCCAGGCTG	3900
40	TGGCCACCT GGTGATGGCC CTTTGTCTCC TGGCAGCCTG AGGCACAGCT GCTGTATG	3960
	TCCTCATCTG TTCTGACTGA AGGATGGAGG TGCTGAATAA ATTAGGCCTC AGGCTCTAC	4020
	CACCAGAGAG CTGGAGAATG GTTCCACGTC ATTCAAGGAC CTGAATTTTT TATGCTCAGG	4080
45	AGCATTGGAA TCCTCTTCTT CCAGGGAGGA ATTAGCCTGC AAGGTTAGGA CTTGAAGAGG	4140
	GAAGGTATTT AATAACTGGG CGAGGATGGG TGTGTTGGCT CACACCTGTA ATCCCAGCAT	4200
50	TTTGGGAGGC TGAGGTGGCC AGATCCCAAG GTCAGAAGAT CGAGACCATC CTGCCTAACA	4260
	TGGTGAACCC CCATCTCTAC TAAAAATACA AAATTAAATT GGCCGGGCGT GAA	4313

55

(2) INFORMATION FOR SEQ ID NO: 147:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1183 base pairs

(B) TYPE: nucleic acid

60

398

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5 GGCAGAGCCT CAAGCTGACT TGGATTATGT GGTCCCTCAA ATCTACCGAC ACATGCAGGA 60
GGAGTTCGGG GGGCGGTTAG AGAGGACCAA ATCTCAGGGT CCGCTGACTG TGGCTGCTTA 120
10 TCAKWWGGGG AGTGCTACT CAGCTGCTAT GGTACAGCC CTCACCTGT TGGCCTTCCC 180
ACTTCTGCTG TTGCATGCGG AGCGCATCAG CCTGTGTTC CTGCTTCTGT TTCTGCAGAG 240
15 CTTCTTCTC CTACATCTGC TTGCTGCTGG GATACCCGTC ACCACCCCTG GTCTTTTAC 300
TGTGCCATGG CAGGCAGTCT CGGCTTGGGC CCTCATGGCC ACACAGACCT TCTACTCCAC 360
AGGCCACCAG CCTGTCTTTC CAGCCATCCA TTGGCATGCA GCCTTCGTGG GATTCCCAGA 420
20 GGGTCATGSC TCCTGTACTT GGCTGCCTGC TTTGCTAGTG GGAGCCAACA CCTTTGCCCTC 480
CCACCTCCTC TTTGCAGTAG GTTGCCCACT GCTCCTCCTC TGGCCTTTCC TGTGTGAGAG 540
TCAAGGGCTG CGGAAGAGAC AGCAGCCCCC AGGGAATGAA GCTGATGCCA GAGTCAGACC 600
25 CGAGGAGGAA GAGSAGCCAC TGATGGAGAT CGGGCTCCGG GATGCGCCTC AGCACTTCTA 660
TGCAGCACTG CTGCAGCTGG GCCTCAAGTA CCTCTTATC CTTGGTATTC AGATTCTGGC 720
30 CTGTGCCTTG GCAGCCTCCA TCCTTCGCAG GCATCTCATG GTCTGGAAG TGTTCGCCCC 780
TAAGTTCATA TTTGAGGCTG TGGGCTTCAT TGTGAGCAGC GTGGGACTTC TCCTGGGCAT 840
AGCTTTGGTG ATGAGAGTGG ATGGTGCTGT GAGCTCCTGG TTCAGGCAGC TATTTCTGGC 900
33 CCAGCAGAGG TAGCCTAGTC TGTGATTACT GGCCTTGGC TACAGAGAGT GCTGGAGAAC 960
AGTGTAGCCT GGCCTGTACA GGTACTGGAT GATCTGCAAG ACAGGCTCAG CCATACTCTT 1020
40 ACTATCATGC AGCCAGGGGC CCGTGACATC TANGACTTCA TTATTCWATR ATTGAGACC 1080
ACAGTGGAGT ATGATCCCTA ACTCCTGATT TGGATGCATC TGAGGGACAA GGGGGKCGGT 1140
45 STCCGAAGTG GAATAAAATA GCGGGCGCTG GTGACTTGCA CCT 1183

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 734 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

GAATTCGGCA GAGTGAAGCA TTAGAATGAT TCCACACTG CTCTTCTGCA CCATGAGACC 60

60

AATCCAGGGC AAGATCCCAT CCCATCACAT CAGCCTACCT CCCTCCTGGC TGCTGGCCAK 120
 GATGTCCCA GCATTACCTT CCACTGCCCTT TCTCCCTGGG AAGCAGCACA GCTGAGACTG 180
 5 GGCACCAGGC CACCTCTGTT GGGACCCACA GGAAGAGTG TGGCAGCAAC TGCTGGCTG 240
 ACCTTTCTAT CTCTCTAGG CTCAGGTA CTCTCCTCAT GCCCATGGYT GGGCCGTGGG 300
 GAGAAGAAGC TCTCATACGC CTTCCSACTC CTTCTGTTTT ATAGGACTTC ACTCCCTAGC 360
 10 CAACAGGAGA GGAGGCCTCC TGGGGTTTCC CCRGGCAGT AGGTCAAACG ACCTCATCAC 420
 AGTCTTCCTT CCTCTTCAAG CGTTTCATGT TGAACACAGC TCTCTCCRCT CCCTTGTGAT 480
 15 TTCTGAGGGT CACCCTAGCC ARCCTCAGGC AACATAGAGA GCCTCCTGTT CTCTCTATGC 540
 TTGCTCTGAC TGAGCCTAAA GTTGAGAAAA TGGGTGCCAA GGCCTAGTCC AGTGTCTTGG 600
 GGCCCTTTTG GCTCTCCCTC ACTCTCTGAG GCTCCAGCTG CTCCTGGGAC ATGCAGCCAG 660
 20 GACTGTGAGT CTGGGCASGT CCAAGGCCTG CACCTTCAAG AAGTGAATA AATGTGGCCT 720
 TTGCTTCTAT TTAA 734

(2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

GGCACAGTGG ACCCCAGACT CCCTCTCCGC CTTTCTCTGC CTGGGGAGAC CCACTGTGTG 60
 40 CATGGCATCA CTGACTCCCA TACCTCTGGC TATCAAAGGT TTCTGCCATG GCCACCCTGG 120
 AAGSAAACCA GAGGGAGGTA GACAGGGAGA TCAGTCCCT TCTACTCTGG TTCCTGCTCT 180
 GTGAAATTGT CTCAGGCTGG CTGTGTCCAG ARGGTCCCTG GTTCTCTCAR GGATGCCAAA 240
 45 TCTACAAGAA TCTCTCCTCT TCCAGTTCCT ATAACCTCTC CTTCCCTTTG TCTCTTTAGA 300
 CCTTGGAGTA GTAGCAGCCA GGTTCCTTCT ATCTCTGGGT TAGTGCAATTA TCTCTGGTGG 360
 50 CTCCCTTACC CAGGACTTTG GGAATGGTCT TTTTGTAAATA CATTCTCCTC AAATAATTCA 420
 ATTTTGAGTG TTCTGTATGT ATCCTGCTGG GAGGTTGTTA TATACAAATC ACTGTGCCCC 480
 55 TTTAGCAGAG AAGGAGACTG AAGCTCAGGG AGGTTAAGTG TCTTTCTCTA GGTGCTATG 540
 TGGAGAAAGT GGCTGACTGG GGAATTGAAT GAGGTCCCTA GTTTCATGCT CCGAGGGCAA 600
 AGANGAATGT CCAATTGGCC TGAGATAAGC CTCTGGTAAA ATGTACTGTA CATAATAGGT 660
 60 AATCAATAAA TGTGGCTGA TGACAAACAT GTTTTCTTTG TTCATTAGTT ATAGTGATTA 720

400

5 TGTTCTAAAT AACTCCMACA AGGAARTCAG CACATTGGA ATATCAWTAT CTTTCATGA 730
 TAATATCTTT CCMYGGAAAG AWAATGATAT TCCMACTGG GAGTGTCCCN ACCARATCTG 840
 ANTCTGTGTA TTGGCCCTGG GGTGGGCCAG CCCCTTAGAC TCTATGGTCT CATTCTCTTT 900
 GTTTACAAAA TTGAGATAAG GCCTTATTCT CTCCCCACCC CACCCATCCA TATTGTTTTG 960
 10 AGAATAAAAT GAGAGGATGT GTGTCAAGGG TGTATTTTGG CAATAGTCTC TGAGCCATTT 1020
 TCTGAGCACC TCCATACTGT TGACACTCAA GTAATATTTC ATCAGCATTC CATTACAGGT 1080
 CCTCCCTTAA TGAGGTGTGC GATGTACAAG AGTGTGAGG TGGCAAAGGA TGGGCTCCTG 1140
 15 AGGAAACACT TAGGAACTG GCCTTTCTGC CATTAAAAGA GACAAACCTT TGTGTGACC 1200
 TAATTAAAGT TTTTAAATT CAATTTGGAA AGTTAGCAAG CTAGCTCCTT TCCAGGWAAA 1260
 20 ATAAGGAGTC AGTGCATGAC CTAACCGGTC CCGGGCTGCT TGCCATTCCA AACAACCTGA 1320
 GTAAGTTTAT CACNTTCTTT CAGGGACTGA GGTTTCCAGG CACAGACTTG GATAAGGAAG 1380
 25 GATGTCCTAT GGGGTCACAT TGATG 1405

(2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2890 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

40 TTATATGCTA CAGCTACAGT AATTTCTTCT CCAAGCACAG AGGANTTTT CTAGGATCAG 60
 GGGGATCGCG CGTCACTTGA TGCTGCTGAC AGTGTCTGTG GGAGCTGGAC GTCATGCTCA 120
 AGTGGCTCCC ATGATAATAT ACAGACGATC CAGCACCAGA GAAGCTGGGA GACTCTTCCA 180
 45 TTCGGGCATA CTCACCTTGA TTATTACAGG GATCCTGCAG GTTATGGGC ATCAAGCAGC 240
 CATATGGACC AAATTATGTT TTCTGATCAT AGCACAAAGT ATAACAGGCA AATCAAAAGT 300
 AGAGAGAGCC TTGAACAAGC CCAGTCCCGA GCAAGCTGGG CGTCTTCCAC AGGTTACTGG 360
 50 GGAGAAGACT CAGAAGGTGA CACAGGCACA ATAAAGCGGA GGGGTGGAAA GATGTTTTCC 420
 ATTGAAGCCG AAAGCAGTAG CCTAACGTCT GTGACTACGG AAGAAACCAA GCCTGTCCCC 480
 55 ATGCCTGCCC ACATAGCTGT GGCATCAAGT ACTACAAAGG GCCTCATGCC ACGAAAGGAG 540
 GGCAGGTATC GAGAGCCCCC GCCCACCCTT CCGGGCTACA TTGGAATTCC CATTACTGAC 600
 60 TTTCCAGAAG GGCACCTCCA TCCAGCCAGG AAACCGCCGG ACTACAACTT GGCCTTTCAG 660

	AGATCGCGGA TGGTCGCCAG ATCCTCCGAC ACAGCTGGGC CTTCATCCGT ACAGCAGCCA	720
	CATGGGCATC CCACCAGCAG CAGCCCTGTG AACAAACCTC AGTGGCATAA AYCGAACCAG	780
5	TCTGACCCGC GCCTCGCCGC YTATCAGTCC CRAGGGTTTT CCACCGAGGA GGATGAAGAT	840
	GAACAAGTTT CTGCTGTTTG AGGCACAGAC TTTTCTGGAA GCAGAGCGAG CCACCTGAAA	900
10	GGAGAGCACA AGAAGACGTC CTGAGCATTC GAGCCTTGGA ACTCACATTC TGAGGACGGT	960
	GGACCAGTTT GCCTCCTTCC CTGCCTTAAA AGCAGCATGG GGTTCCTTCT CCGCTTCTTC	1020
	CTTTCGCCCT TGCATGTGAA ATACTGTGAA GAAATGCCCC TGGCACTTTT CAGACTTTGT	1080
15	TGCTTGAAAT GCACAGTGCA GCAATCTTCG AGCTCCCACT GTTGCTGCCT GCCACATCAC	1140
	ACAGTATCAT TCCAAATTC AAGATCATCA CAACAAGATG ATTCACCTCG GCTGCACCTC	1200
20	TCAATGCCTG GAAGGATTTT TTTAATCTT CTTTTAGAT TTCAATCCAG TCCTAGCACT	1260
	TGATCTCATT GGGATAATGA GAAAAGCTAG CCATTGAACT ACTTGGGGCC TTAAACCCAC	1320
	CAAGGAAGAC AAAGAAAAC AATGAAATCC TTTGAGTACA GTGCTTGTCG ACTTGTTTAC	1380
25	AATGTCTCC TTTTAAAAA AAAAAATGA GTTTAAAGAT TTGTTCAGA GAGTAAATAT	1440
	ATATCCATTT AATGATTACA GTATTATTTT AAACCTTAAG TAGGGTTGCC AGCCTGGTTT	1500
30	CTGAAAAACC AAATATGCCG GACAGGGTGT GCCCACACCA AGAAGACGGG AAGACCTGGC	1560
	TTGTGACCCG GCCTTCCCAT GTCTTCTGG TCTCAGCCGC GAAGTGCCCT ATCCTGGAAG	1620
	TATGAAATGT TAGCCAATTA ATACCAAGAC ACCTCATCTG CTCCTTCCCC AGTGCATGGG	1680
35	GTTCCTCTGT AAAACTGTTT GCACATGGCC AGGGGAGGGA ACTAGGACCC TTGTGTCTTG	1740
	TCTGAGCCTT ATGGAGGCAG GACCGTGTCA TTGGCGGATG TGTCTGCTC CATGAGATG	1800
40	GATGGCAAAC CCCATTTTTA AGTTATATTT CTTTGATTTT TGTAAATTA GAGGTGTAGG	1860
	TTTTGTTTTT TGTTTTTTTG TTTTTTTTAA AGAGAAACAT TTATAACTGG ATAGCATTGC	1920
	AGTGAAAGCA CCTTGGGATG TTGGAGCTAA TGCCAGCTGT TTATACTGCT CTTTCAAGAC	1980
45	AGCCTCCCTT TATGGAATTG GCATTAGGGA ATAAACAAGC CTTTAAACGT GATAAAAGAT	2040
	CAAAAACCTG GTTAGACATG CCAGCCTTTG CAAGGCAGGT TAGTCACCAA AGACTAACCT	2100
50	CCAAGTGGCT TTATGGACGC TGCAATATAGA GAAGGCCTAA GTGTAGCAAC CATCTGCTCA	2160
	CAGCTGCTAT TAACCTATA ATGACTGAAA TGACCCCTCC ACTCTATTTT TGTGTGTTTT	2220
	TGCACAGACT CCGGAAAGT GAAGCTGCC AATCTGACTA GTACTCAAAT GTGAGGAAT	2280
55	GCTGGTCTTG GATTTTTTTT CCATTAAATT CAGCTGATCA TATTGATCAG TAGATAAACC	2340
	TAAATAGCTT CAAATTTTAA AAGTGAATT GCAGTGTTTT TTCACTGTAT CAAACAATGT	2400
60	CAGTGCTTTA TTTAATAATT CTCTCTGTA TCATGGCATT TGTCTACTTG CTTATTACAT	2460

TGTCAATTAT GCATTTGTAA TTTTACATGT AATATGCATT ATTTGCCAGT TTTATTATAT 2520
 AGGCTATGGA COTCTGTGTC ATATAGAAAG ACAGAAATCT AGCTCTACCA CAAGTTGCAC 2580
 5 AAATGTTATC TAAGCATTAA GTAATTGTAG AACATAGGAC TGCTAATCTC AGTTGGCTCT 2640
 GTGATGTCAA GTGCAGAATG TACAATTAAC TGGTGATTTC CTCATACTTT TGATACTACT 2700
 TGTACCTGTA TGTCTTTTAG AAAGACATTG GTGGAGTCTG TATCCCTTTT GTATTTTAA 2750
 10 TACAATAATT GTACATATTG GTATATTTT TGTGAAGAT GGTAGAAATG TACTATGTTT 2820
 ATGCTTCTAC ATCCAGTTTG TACAAGCTGG AAAATAAATA AATATAACAT AAAAAAAAAA 2880
 15 AAAAAAAAAA 2890

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2399 base pairs
 (B) TYPE: nucleic acid
 25 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

30 GAACCTTTTCC ATCTGGCAAA CCGGAAACTC CATCCCCATT AAACCAACTC CCCCTTTTGG 60
 TTTCCCCCCC AGNGGAATAG AATTGGACN CCCATATAAA TCCAGGAAAC CACCTAAATT 120
 CTTTAGTNGT TTGTGTTTGC AAGATCTAAG GTCATGGTAA ACATTAAGTT CTAAAAATT 130
 35 TTGGGAGGGA CCACTGCACC TCTCCCTCTG AATTGTTTNC CAATTTAAAA TTGGAGTAAG 240
 GTTTTAAAAAT GTCTNATTCC ATTGGAAGGG TMTGTTATTT CATTTTGAGC CCAGAGGGGA 300
 40 GAGGCACATT TAAATATCA GAATTAGATT AGCTTTGAGT TTGTACAATT CGGAACATAA 360
 TAGATTTTCA TAAATTATGT GTGCTTGTT GGAAGTGTCA ACTGTCTTTA TGTCTGCTTG 420
 TAAAAGTTTC AAAATATGTT TTCCCTCAAA AAGGCAACGT TACTTCATTT GCTTGAATAT 480
 45 TATGATAGGA ATGCTTACTG ATATTACTTG ATAGTCATAT ATAGCCTAGG AAATTTAACA 540
 TATATATAAC TATAGCAGTA TTAATAATGA TAGTTGTACT TCTTTAAAAC ATTAAATTG 600
 50 AGGAAACTTT AATGCTGTCT CGTGTACATT GCTTTACTAC AGTGAGGGGG AATATCCTTT 660
 AGATTGAGCC TCAATTTACT GGTAGTAGT ATGTGAACTC TGGTATAAAA AGGTAACTA 720
 GACAGTAGAG CCGATGAATT AAAATTGTAA ATTGCTACAT TGGCATTTTC TACCTCCTTT 780
 55 TCTGTCAGAG TATTACTTTT TCCAGCATTT ATTCTTATTT GTGAGTAAAG AGGAAATGGG 840
 AACCTGAGGT TAAAATTGAC ATTTTTGTTT CATTGAGAAT TTAAGCAGTA GGTACAGGAG 900
 60 AAGTGACTTG TCACATTAAT TGGTGCTTA AATGTGTAAC TACAAGTTGT GATCGACATG 960

	TACAAAATGT CTAAGAAAGG TCATATGCTG AATATTTTAC TTTTCCTGTA TAGTCTGCAT	1020
5	GATTGTGTTT ATAAACCCAG CTTATTTTCT CCAAAAAGCA AAATGGTCTT GTAATTTTTA	1080
	AAGTAAATA AACGTGCCAT TTTGTCTGCA ATCTATAATT TCAGGAAGTT ATTGRAAGTT	1140
	CTGACTCAGG GCTTTTAAAC AGTTCAAGCA ATTGTCAATT ATATTTTGGT AACTCCATCT	1200
10	GTGTAATTCT CCAGTGCCCTT GAAAGAATTA TTAAGTTGGC AACACTATTA AACTTTTATA	1260
	AAAGATGGTC TTTAGTGCAC GTGTATCATT ATATACAGCT TTAAAGTCA TATTGCTTAG	1320
15	CTTGTTAATA ATGATTCTGC ATGTGTGCTG GGTTCGGTA ATTCTTTAAA GGAAGTTTTC	1380
	TAGATTTGCA CTTGATGTTT GTTTTTTAAA AACTGATTAT TTATGGCCGT GACACTGTTA	1440
	CCAGAAAAGT AATTCTAATT AAGTTATTAT GCAAAGTCAT CTATAAGTAG CATCTGGGAA	1500
20	GAGGAGATSG AGGCCACAGT TTGCTATTTT ACTATGAAAG GAGGATCTGT TTGGCAACA	1560
	TAGATGTCTT TCCCTTCAAA TGAGGGGAAA AAAAAAGACC CTTGTTCAT ATGGATTCTG	1620
25	TTGTAAAAAA TTATTTTAA AGGAAATCAC AAATTGTATG TCATTCTTAA TGCTAGTCTT	1680
	ATAGAATAAA TCCATAAAAT TGTTTTTATG TTCAGTATGT TTATGTCATT CTAAATGCAG	1740
	CAAATTCAT GATAGCAGTT CAATTGACTC ATAGCAGTGT TTTGTATTTT TTCTAATTCT	1800
30	TTAGCTTTCA ATATTGGATT AAAGTCTTGT TTGTGAATAT AGTTTCCGTA TGGCAATGA	1860
	TTTCTTGCTT ATTAGCTTTT GTTAAAGAAT GCTTAGTAAG AGCTAAGCTT TTAAGAATA	1920
35	TGCAACATT TATCGTAAAT AAAACCTATG GTGTAATATC ATATAATGCT TTTCTTTGAT	1980
	CTTTGGAGAA TTATTCTTTT ATAGTAGTAT ACATGAATTT TGATTTTAA AGCATTTAAA	2040
	AACAAATCTC AATACATTAA AAAACCTGTT ATTGTTAAAA RGGAAATAC CATGCCCTTA	2100
40	AGAAACAAGG ATGTACATCT TCAATTCAGC ATRAGTGTCC ACATCTAGAA GGCTCTCAT	2160
	GCAGTTGTTT ACAGTTAAGG TACCTCTATC TAAAGGGCCA AAGAAGCATT TCATATTTTA	2220
45	ACACCTCACA TTCTTTCAGG ATTAAGACAT ATGAAAATAG TCTGAATAGG ATAAATTTGG	2280
	ATAGGAAGTA ACTTAACCAG TCTGGGAAGA TTCAGGCTTT TTCTATKAAA AAGCTTATTC	2340
50	CTCTTCACAA CTCNGGTGGT AGGNTTTCAT TTTCAAGAG GGTAGATATT TTAAGCCA	2399

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

	CGTGCCTGTA GTAAGCTCAT CCCTGCCTTT GAGATGGTGA TGCCTGCCAA CGACAATGTT	60
5	TACCACCTGG ACTGCTTTGC ATGTCAGCTT TGTAAATCAGA GATTNTGTGT TGGAGACAAA	120
	TTTTTCCTAA AGAATAACWT GAYCCTTTGC CARACGGACT ACCAGGAAGG TTTAATGAAA	180
10	GAAGGTEATG CACCCCGGT TCGCTGATCT ATCAACATCA CCCCATTAA G AATACAAAGC	240
	ACTACATTCT TTTATCTTTT TTGCTCCACA TGTACATAAG AATTGACACA GGAACCTACT	300
	GAATAGCGTA GATATAGGAA GGCAGGATGG TTATATGGAA TAAAAGCGG ACTGCATCTG	360
15	TATGTAGTGA AATTGCCCA GTTCAGAGTT GAATGTTTAT TATTAAAGAA AAAAGTAATG	420
	TACATATGGC TGGATTTTTT TGCTTGCTAT TCGTTTTTGT GTCACTTGGC ATGAGATGTT	480
20	TATTTTGGAC TATTGTATAT AATGTATTGT AATATTTGAA GCACAAATGT AATACAGTTT	540
	TATTGTGTTA CCATTGTGT TCCATTGCT YCTTTGTATT GTTGCATTTA GTACAATCAG	600
	TGTTTAAACT TACTGTATAT TATGCTTTC TGTATTTACC AGCTATTTTA AATGAGCTGT	660
25	AACCTTCTAG TAAAGAATTG AAAAGCAAAT CCTCACTAAA GGATACACAG GATAGGATAA	720
	AGCCAAGTCN CATCAACATT AAAAAATACT AAAANANAAA ACACAAAAAA AAAAAANCCC	780
30	GGGGGGGGCC CGGAACCCAT TC	802

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 461 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

45	CTAGGAGCAC CGAGCAGCTT GGCTAAAAGT AAGGGTGTGG TGCTGATGGC CCTGTGGCGA	60
	CTGACCCGGG CTCTGCTCTC TGTGAACCTG GCGCCCCCGA CCGTCGCGGC CCCTGCCCGG	120
	AGTGTGTTCC CCGCCGCCCA GATGATGAAC AATGGCCTCC TCCAACAGCC CTCTGCCTTG	180
50	ATGTTGCTCC CCTGCCGCCC AGTTCTTACT TCTGTGCCCC TTAATGCCAA CTMTGTGTCC	240
	TGGAAGAGTC GTACCAAGTA CACCAATTACA CCACTGAAGA TGAGGAAGTC TGGGGGCCGA	300
55	GACCACACAG GTGGGAACAA GCACAGGGGG ATTTAAGCAG TCAAAAGGAA AAACATGTTA	360
	AGACCCTAGA CTTGTATATT GACACACTTG TACCTTGTA GGCAGAGGAA TGTAATTAAA	420
60	AAGCACTTAT TTGGCAGGAA AAAAAAAAAA AAAAAAAAAA C	461

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2388 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

	CCCCACGCGT CCGAAAGCGG AGAAGCGTCG TGGCCCTGTT GTGGAGTACG CTTTGGACTG	60
15	AGAAGCATCG AGCCTATACG ACCCAGCTGT TGCCATGACG GCCCAGGGGG GCTGGTGGCT	120
	AACCGAGGCC GCGCCTTCAA GTGGGCCATT GAGCTAAGCG GGCTTGGAGG AGGCAGCAGG	180
20	GGTCGAAGTG ACCGGGGCAG TGGCCAGGGA GACTCGCTCT ACCCAGTCGG TTACTTGGAC	240
	AAGCAAGTGC CTGATACCAG CGTGCAAGAG ACAGACCCGA TCCTGGTGGG GAACCGCTGC	300
	TGGGACATCG CTTTGGGTCC CTTCAACAG ATTCCCATGA ATCTCTTCAT CATGTACATG	360
25	GCAGGCAATA CTATCTCCAT CTTCCCTACT ATGATGGTGT GTATGATGGC CTGGCGACCC	420
	ATTCAGGCAC TTATGGCCAT TTCAGCCACT TTCAAGATGT TAGAAAGTTC AAGCCAGAAG	480
	TTTCTTCAGG GTTTGGTCTA TCTCATTTGG AACCTGATGG GTTTGGCATT GGCTGTTTAC	540
30	AAGTGCCAGT CCATGGGACT GTTACCTACA CATGCATCGG ATTGGTTAGC CTTCAATTGAG	600
	CCCCCTGAGA GAATGGAGTT CAGTGGTGGG GGACTGCTTT TGTGAACATG AGAAAGCAGC	660
35	GGCTGGTCCC TAGTATTG TGCTCTATT ACATCCTTCT TTAAGCCCGAG TGGCTCCTCA	720
	GCATACTCTT AAACATAACA CTTATGTAA AAAGAACCAA AAGACTCTTT TCTCCATGGT	780
	GGGGTGACAG GTCCTAGAAG GACAATGTGC ATATTACGAC AAACACAAAG AAATATACC	840
40	ATAACCCGAG GCTGAAAATA ATGTAGAAAA CTTTATTTTT GTTTCCAGTA CAGAGCAAAA	900
	CAACAACAAA AAAACATAAC TATGTAAACA AGAGAATAAC TGCTGCTAAA TCAAGAAGTG	960
45	TTGCAGCATC TCCTTTCAAT AATTAAATG GTTGAGAACA ATGCATAAAA AAAGTTGCAC	1020
	AAGTTCCTTA TTTTCTTAA TATTTCACTT CTATTTAATA CAAGCTGGGA CATAAAAAAT	1080
	CTGTTGGGGA TACCTGGGGG AAGATGTGAG AAACAAATGC TGAATTCAGC TTATACATGA	1140
50	TGAAAAGAAA AACCAGACAA AAGGAGCACA TAAATATGCA TACAGTGTA CTGTTATTAT	1200
	TTTAATACCC ACCATAAGGG ATTTTGTGA GCATGTTTAG GGGGAACGAG GATTGGTGGG	1260
55	ATCCTTGGGG CCACAGGAAT CTGAGGCAAC GGAAGATATA TAGAGTGATC GTCCCCCTGC	1320
	CGAAGGAACC TGGCAYCTGT CAAGCAGATG CTGCAGTTCA AACTTCAGCT TTTAAGATAG	1380
60	ATAGCTATTG AAGGCAGAGG GTCAGCAGGA GGATGTGTAT TTCTAATCTA CCGTGGTAAA	1440

5 GTCATAGGTA AGACTCAAAA GCGGGATCTT ATTCAAAAGG CAGGTATTTC CTTTGTTTTC 1500
 TGTCTTGAAA TAGCCCCCTC CCTTAAGGTG CATCTCTCTCA AGTTTTCTAGT ATTGCTTTAT 1560
 10 TTGCAGTGAT TAAAGAGAT GAGAGACTTT GGAGACAGAC AACGTAAGCA ACACATACAC 1620
 ACATGAAATA CTCTAGACAG AGATGAATAT AAATCTGGCC TAATPACCAG TTTTCCATGT 1580
 AACAGTGATT TTGTGTTTCG GCGTGAAGCA GTGGTTATAT TAAAAGCCAC TAATTCCTTT 1740
 ATCCCTTTAA AAGATTTTCA CAATTCTCCA ACCACAAACA GCACCTTCTAA AACTAACTTT 1800
 ACTTTCTGCC CATAATTTGT TCTACATGGA AAAAAAAAT ATTACTTTTCG CCAGGGGTGT 1860
 15 GTGTAAATGT GGCAGAATTC CTAGGCAGGC TGACCTTTAC AGTATGGGCC TTTAAGATAC 1920
 TGGATCCTGG TTGGGCAACA AGTGTACGC CTGAAGTTTC TGAAAACAAA TTAGAAGACT 1980
 GTTGGCTTGG CTAATCTCGT AGTTCAGGCC CAAGTTTCTG TAGTCAGAAT GAAGAATAAA 2040
 20 ATTGAAAGAA AAAGGGGGAA ATGCTTATAC TTGGCATTAA GTTGAATGCC TCAAGTCTTA 2100
 ACTATGGCTT TGTAGATGAG GCAAAAGATT TCTTAGTGGT AAAATTTCTT CAACAGGTCA 2160
 25 ATGCCAATCT GTATGCCATT TTAGTAAAGT AGGTAAGGAG AGTAGCGGCT CAGTAACTTT 2220
 GGCCTAAAG AAAGAGTGTG GCTCTAGAAC TTCCAATCCC ATTGCTAGAT GTGCCCTTTA 2280
 AAAGATGGTC CAGTGCTTTC AGGGAAGGAT GTTTAGCCAG TTTTCCTAGT ATTTGTTCTT 2340
 30 TAAGATTTTT TGACCTGTGC TTAATAAGAC GGACGGGTGG GTCCAGCC 2388

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 642 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

45 AAAACAGACC ATTAAAAAC TCAGACAAGA TTATATTAA TATATTAAAT ACTAAAAAGG 60
 CACAAGATTA CACTGAACAT ATTAGCTACT AAAAAGGCAC TGCTAAGACA TTCAAGCAAA 120
 50 TAGCTATTAC ACACTACTGC AGATTTTACA GGTTCCTAAT TCTAACATAT GTTTGAAAAA 180
 TCCGTGAGTA TTCCAAATA TATTTAATAA TGGAATATCT GCATTAATAT ACCATCCATG 240
 55 TGTTTTACC ATTGCTTA ATATTGAATA TACTGTTTAC CTCACACTPA AAAGAAAACC 300
 AGAAGCTTA TTTGTGATTT TGGGACTGGA AGCTTCCATT TTTGTGTCAA AAATGAATCC 360
 TGATTCTTAT GGAAATCTCT GTTATTAAGA TATTTCAACA TGAGACAACA CTGAAGATCA 420
 60 AATTGGTTTT AGTATCACTA TCTTCTCTCC TCGTTTCTCT CTTACTCCTC ATCCTCCCAG 480

AATCTACCAG TTTATGGTAG AAAGATGGGA ACCTTATTTG AATGTGTTTT TTTTTTTCCA 540
 TGATGTCCAA TTTTGTGTG CGAAAGGATT TGCATAAAAT TTTTGTTTAA ATTTTGGTAG 600
 ATTTTTATCT ATACAAATTT AAATAAAATT ATGTTTTGTA AG 642

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1251 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

GCGGCTGCCC CTCCACGGAG TGGCTGATCA TGTGGGCTGT GATCCACAAA CCGGTTCTT 60
 TGTCCCTCCT AATATCAAAC AGTGGATTGC CTTGCTGCAG AGGGGAAACT GCACGTTTAA 120
 AGAGAAAATA TCACGGGCGG CTTTCCACAA TGCAGTTGCT GTAGTCATCT ACAATAATTA 180
 ATCCAAAGAG GAGCCAGTTA CCATGACTCA TCCAGGCACT GAGCATATTA TTGCTGTCTAT 240
 GATAACAGAA TTGAGGGGTA AGGATATTTT GAGTTATCTG GAGAAAAACA TCTCTGTACA 300
 AATGACAATA GCTGTTGGAA CTCGAATGCC ACCGAAGAAC TTCAGCCGTG GCTCTCTAGT 360
 CTTCTGTGTA ATATCCTTTA TTGTTTGTAT GATTATTTCT TCAGCATGGC TCATATTCTA 420
 CTTCAATCAG AAGATCAGGT ACACAAATGC ACCCGACAGG AACCAGCGTC GTCTCGGAGA 480
 TGCAGCCAAG AAAGCCATCA GTAAATGAC AACCAGGACA GTAAAGAAGG GTGACAAGGA 540
 AACTGACCCA GACTTTGATC ATTGTGCACT CTGCATAGAG AGCTATAAGC AGAATGATGT 600
 CGTCCGAATT CTCCCTTGCA AGCATGTTTT CCACAAATCC TGGGTGGATC CCTGGCTTAG 660
 TGAACATTGT ACCTGTCCCTA TGTGCAAACT TAATATATTC AAGGCCCTGG GAATTGTGCC 720
 GAATTGCGCA TGTACTGATA ACGTAGCATT CGATATGGAA AGGCTCACCA GAACCCAAGC 780
 TGTTAACCGA AGATCAGCCC TGGCGGACCT CGCCGGCGAC AACTCCCTTG GCCTTGAGCC 840
 ACTTCGAACT TCGGGGATCT CACCTCTTCC TCAGGATGGG GAGCTCACTC CGAGAACAGG 900
 AGAAATCAAC ATTGCAGTAA CAAAGAATG GTTTATTATT GCCAGTTTTG GCCTCCTCAG 960
 TGCCCTCACA CTCTGCTACA TGATCATCAG AGCCACAGCT AGCTTGAATG CTAATGAGGT 1020
 AGAATGGTTT TGAAGAAGAA AAAACCTGCT TTCTGACTGA TTTTGCCTTG AAGGAAAAAA 1080
 GAACCTATTT TTGTGCATCA TTTACCAATC ATGCCACACA AGCATTATTT TTTAGTACAT 1140
 TTTATTTTTT CATAAAATTG CTAATGCCAA AGCTTTGTAT TAAAGAAAT AAATAATAAA 1200

ATAAAAAAAAAA AAAAACCCTG GGGGGGGCCC GGTCCCAAT TGGCCCTATG G

1251

5

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2127 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

15

CCGGCGGGAG AGGAAGCTG CAGCGAGAGG CCGGATCTC AGCGCGGGAG CACTGCTTCT 60

GCGGCAGGCC CCTGAGGGAG GGAGCTGTCA GCCAGGGAAA ACCGAGAACA CCATCACCAT 120

20

GACAACCAGT CACCAGCCTC AGGACAGATA CAAAGCTGTC TGGCTTATCT TCTTCATGCT 180

GGGTCTGGGA ACGCTGCTCC CGTGAATTTT TTTCATGACG GCCACTCAGT ATTTCACAAA 240

CCGCCTGGAC ATGTCCGAGA ATGTGCTCTT GGTCACTGCT GAACTGAGCA AGGACGCCCA 300

25

GGCGTGAGCG CNCCCTGCAG CACCCTTGCC TGAGCGGAAC TCTCTCAGTG CCATCTTCAA 360

CAATGTCATG ACCCTATGTG CCATGCTGCC CCTGCTGTTA TTCACCTACC TCAACTCCTT 420

30

CCTGCATCAG AGGATCCCCC AGTCCGTACG GATCCTGGGC AGCCTGGTGG CCATCCTGCT 480

GGTGTCTCTG ATCACTGCCA TCCTGGTGAA GGTGCAGCTG GATGCTCTGC CCTTCTTTGT 540

35

CATCACCATG ATCAAGATCG TGCTCATTA TTTCTTTGGT GCCATCTGCG AGGGCAGCCT 600

GTTTGGTCTG GCTGGCCTTC TGCCTGCCAG CTRACACGGC CCCATCATG AGTGGCCAGG 660

GCCTAGCAGG CTTCTTTTGGC TCCGTGGCCA TGATCTGGCG TATTGCCAGT GGCTCGGAGC 720

40

TATCAGAAAG TGCCTTCGGC TACTTTATCA CAGCCTGTGC TGTRATCAT TGGACCATCA 780

TCTGTTACCT GGGCCTGCCC CGCCTGGAAT TCTACCGCTA CTACCAGCAG CTCAAGCTTG 840

AAGGACCCCG GGAGCAGGAG ACCAAGTTGG ACCTCATTAG CAAAGGAGAG GAGCCAAGAG 900

45

CAGGCAAAGA GGAATCTGGA GTTTCAGTCT CCAACTCTCA GCGCACCAAT GAAAGCCACT 960

CTATCAAAGC CATCCTGAAA AATATCTCAG TCCTGGCTTT CTCTGTCTGC TTCATCTTCA 1020

50

CTATCACCAT TGGGATGTTT CCAGCCGTGA CTGTTGAGGT CAAGTCCAGC ATCGCAGGCA 1080

GCAGCACCTG GGAACGTTAC TTCACTCTG TGCTCTGTTT CTTGACTTTC AATATCTTTG 1140

55

ACTGGTTGGG CCGGACGCTC ACAGCTGTAT TCATGTGGCC TGGAAGGAC AGCCGCTGGC 1200

TGCCAAGCTG GNTGCTGGCC CGGCTGGTGT TTGTGCCACT GCTGCTGCTG TGCAACATTA 1260

AGCCCCGCGG CTACCTGACT GTGGTCTTCG AGCAGGATGC CTGGTTCATC TTCTTCATGG 1320

60

CTGCCTTTGC CTTCTCCAAC GGCTACCTCG CCAGCCTCTG CATGTGCTTC GGGCCCAAGA 1380

409

AAGTGAAGCC AGCTGAGGCA GAGACCCGAG AGGCATCATG GCCTTCCTCC TGTGTCTGGG 1440
 TCTGGCACTG GGGGCTGTTT TCTCCTTCCT GTTCCGGGCA ATTGTGTGAC AAAGGATGGA 1500
 5 CAGAAGGACT GCCTGCCTCC CTCCTGTCTT GCCTCCTGCC CCTTCCTTCT GCCAGGGGTG 1560
 ATCCTGAGTG GTCTGGCGGT TTTTCTTCT AACTGACTTC TGCTTTCCAC GGCCTGTGCT 1620
 10 GGGCCCGGAT CTCAGGCCCC TGGGAGGGA GCCTCTGGAC GGACAGTGGG GACATTGTGG 1680
 GTTTGGGGCT CAGAGTCGAG GGACGGCGTG TAGCCTCGGC ATTTGCTTGA GTTTCTCCAC 1740
 TCTTGGCTCT GACTGATCCC TGCTGTGCA GGCCAGTGGG GGCTCTTGGG CTGGAGAAC 1800
 15 ACGTGTGTCT CTGTGTATGT GTCTGTGTGT CTGCGTCCGT GTCTGTGAGA CTGTCTGCC 1860
 GTCTGGGGT GGCTAGGAGC TGGGTCTGAC CGTGTATGG TTTGACCTGA TATACTCCAT 1920
 20 TCTCCCTGC GCCTCCTCCT CTGTGTTCTC TCCATGTCCC CCTCCCACT CCCCATGCCC 1980
 AGTTCTTACC CATCATGCAC CCGTACAGT TGCCACGTTA CTGCCCTTTT TAAAAATATA 2040
 TTTGACAGAA ACCAGGTGCC TTCAGAGGCT CTCTGATTTA AATAAACCTT TCTTGTMTT 2100
 25 TTCTCCATGG AAAAAAAAAA AAAAAA 2127

30

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 1625 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

40

45

50

55

60

CAAAAGATCT ATAATCAGGA CATTGTTTAT GTAAGTTGGA CAANAAAAAT TCTTCCCCTT 60
 TATGTCCACC CTTCTATGA TTGCAAGACA AAATTTCCCT CTTTACCTC ATCCCTATAA 120
 CATGGGAGGC TCAGAAAAAT GAGGGGAGAT GGAACCAGAT ACAAGGAGAT CCAATAAGAG 180
 AAGCTTATTT AAATATTGTG AAATAAAGGA AGAMCCAAAG CATTTTTTTA AGTGGGAAT 240
 CCTTTGAAC AGTTATTATT TATCCATATT ATTAAYAACA TCTTTCTGA CAAAATCCAT 300
 CAGATGAAGT GTAAATGGAT AATCTTTTAA TGGATCTAAA CCTAGAAAGT TTCACTTACT 360
 GTTCATGCC GTGTTCCAGA ATTGTGAAAT GGTGTGTGCT TTTGCTTCC AAGTTCTTCT 420
 55 CTGCTCCTC TTAATTCTCT AATCCATGT CTTACAGAAG AATGAGAAAT TTCTTTCTTA 480
 CTTGAGTATC ATGCTCTAAA AACTTGGCT TCAGTCACAG AAACGCTGGC TCTCCTGTGC 540
 TTATATTGAA GGCRACTGCC TTAAATTCTT GGGCCCTCTT ATATTTTAA GGTGCAAAAT 600

410

	TTGAAGTCTC AGTCACCAGA CACAGGTTCT ATACAATTAA TGATGAGCTG GAGAAGTAAT	660
	ATGTAGCTAA TTTTTC AAAA GCATTGAATA TACTTTCCCG AAAGAAAACA GAATTAAAT	720
5	ATTGCCACAT CTGCCCAGAA TCCCATCTGA CACCTTAACT TTGTCAGGTT TCCTACAACT	780
	TGCTAATCAA GTTTTATACA TTCTAAATCT CCCAGTTTC TTTGGGGCTG GAAGATGCCA	840
10	CTTCCATTTA ATAGAACTT TGAAATCTTG GCGTAAGGGA GCAGTGGGGG GACTAGGGAG	900
	AAGGATAAGA AATAGAATTA TTGAAAAGCC CCCACCAGG ACCTTCCTGG CCAQAATATG	960
	CAGAGTAATT CCTGCTGGCT TCACCTTTGA AAGTCCCTCG AACTATGCA GATGAAACTG	1020
15	AGTCTGTTTT TGATATTGTC AGATGTATTC TACCTTGGAA GTCCCNACAC CTAAACTGGA	1080
	ATTCTTGAT TTACATCTCC TCCACTGTCC CCCACACCAC CCTCAATTC CTGCTGGCCC	1140
20	TGCTAATGTT AAGCATTTTT CTCTTGTTAT CATCAGGTTT ACATTAAAPM CAGTACTTA	1200
	CAAACTGACT TGAAGCACAG ATACTTTTAC GAATGTGATA AATATTTTC TTAAGAAAAG	1260
	GAAAGAGGAT GTGGGTCAA TAAAACACCG CATGGATGTT GATTGGTGAA TACTGGTGTA	1320
25	AGAAAAGGGA GCTCAGGAAT TTTTATTACT GTATTTGTAA ATGAGTTTGA AGGAATTTGT	1380
	AAATGCCACT GGTACATTTT TRAGGTGACA CATTTGCTCC TTATAAAGTT ATTAAAAAT	1440
30	ACAGGGTAAG CTTAAATGAC GTTTGCCAGT AGTTTTACTT TATATAATCA ATATTGATAT	1500
	TGTTGCTGAA CTATGTAAT TTATGATGCA TTTTTCAGTC CCTTTCAGA GCAAATGCTT	1560
	TTGCAATGGT AGTAATGTTT AGTTTAAAT GACTTAATAA ATTMTACCT GAGCAAAAAA	1620
35	AAAAA	1625

40 (2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1637 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

50	CGGGGTCACC AGTTATTAGA GGAAGTAACA CAAGGGGATA TGACTGCAGC AGACACATTT	60
	CTGTCCGATC TGCCAAGGGA TGATATCTAT GTGTCAGATG TTGAGGACGA CCGTGATGAC	120
55	ACATCTCTGG ATAGTGACCT GCATCCAGAG GAGCTGGCAG GAGTCAGGGG ACATCAGGGT	180
	CTAAGGGACC AAAAGCGTAT GCGACTTACT GAAGTGCAAG ATGATAAAGA CGAGGAGGAG	240
	GAGGAGAATC CACTGCTGGT ACCACTGGAG GAAAAGGCAG TACTGCAGGA AGAACAGGCC	300
60	AACCTGTGGT TCTCAAAGGG CAGCTTTGCT GCGNATCGAG GACCATGCCG ATGAAGGCCC	360

411

5 TGGAGATCAG TCAGGCCCCAG CTGTTATTTG AGAACCCTGYG GAAGGGACGG CAGCAGCCAC 420
 AGAAGCAGCA GCTGCCACAG ACACCCCTTT CCGTTTGAAG GACTGAGATA ATGTCTCCCC 480
 TGTACCAAGA TGAAGCCCTT AAGGNAACAG AGGCTTCTTC GGGGACAGAA GCTGCCACTG 540
 GCCTTGAAGG GGAAGAAAAG GATGCCATCT CAGACACTGA TAGCAGTACT ACCAATGAGG 600
 10 AAGAAGAGAG CTGGGAACCC TCCGTGGTAA GAAGCGAASC GTGGGCTAA AGTCAGATER 660
 TGACGGGTTT GAGATAGTGC CTATTGAGGA CCCAGCGAAA CATCGGATAC TGGACCCCGA 720
 AGGCTTGGCT CTAGGTGCTG TTATTGCCTC TTCCAAAAG GCCAAGAGAG ACCTCATAGA 780
 15 TAACTCCTTC AACCGGTACA CATTTAATCA GCATGACGGG CAGCTTCCGG ACTGCTTGT 840
 GCAAGAGGAA AAGCAGCACC GGATACGACA GTTGCTGTT GTTAAGAAGG AGGTGGAGCA 900
 20 TTACCGGAAA CGCTGGCGGG AAATCAATGC ACCTCCCATC AAGAAGGTGG CTGAGGCTTA 960
 GGCTAGAAAG AAAACGAGCA TGCTGAAGAG GCTGGAGCAG ACCAGGAAGA AGGCAGAAGC 1020
 CGTGCTGAAC ACAGTGGACA TCTNCAGAAC GAGAGAAAGT GGCACACCTG CGAAGTCTCT 1080
 25 ACAAGAAGGC TGGCTTGGC AAGGAGAAAC GCCATGTCAC CTACGTTGTA GCCAAAAAG 1140
 GTGTGGGCGG CAAAGTGCGC CGGCGAGCTG GAGTCAGAGG TCATTTCAG GTGCTGGACT 1200
 30 CAAGGATGAA GAAGGACCAA AGAGCACAGC AACGTAAGGA ACAAAGAAA AAACACAAAC 1260
 GGAAGTAAGC AGAGCTGCCA GGCTCCAGG AGAGCATGGG GACTAGGAGG AAGGGTGTGG 1320
 CATGGCTCAG TCTGGCCCCC TTGATTACCG GCCTAGCCCC TGCTCACATC ACAGCTGTCT 1380
 35 GAAGAACAGT GAGGTGAGT GCCTAGAACT CCCGTGGTGG TCCTGAGCAG AGAGGAGGAT 1440
 GTCTCTCTGC CTGCTGAAG GTCTCCCATG AAAACACTGC TGAAGTGTGT TGACACTCAT 1500
 40 GACCCTTTTT TTAACCGTT AAGGGAAGT TCGGTGTTGG ACCGATACTC AATGTAGTCA 1560
 GTCTACACCT GGACGTGTGG GCCACTTAAG CCTCCCCAC CCGCATCTTA TTCCTRAATA 1620
 45 AAACCAGGAT AATGGAAPAA AAAAAAAAAA AAAAAAAG GGGGGGCGCN TPAAGGNC 1680
 CANNNTT 1687

50

(2) INFORMATION FOR SEQ ID NO: 160:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1842 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

	GGATGACAGA TTGCGACANA GATTTGTGAC CTTTCCTGCT GAACTTCAGA GGGAGCTGAA	60
	ANCAGCGTAT GATCAAAGAC AAAGGCGGG CGAGAACAGC ACTCACCAGC AGTCAGCCAG	120
5	CGCATCTGTG CCCCCAGAAT CTTTACTTC ATCTAAAGGC AGCAGTGAAA GAAAAGAAAA	180
	GAAACAAGAA GAAAAAACC ATTGTTTAC CAAAAGGAT TCAGAGTCCT TTGAATAACA	240
10	AGCTGCTTAA CAGTCCTGCA AAAACTCTGC CAGGGGCTG TGGCAGTCCC CAGAAGTTAA	300
	TTGATGGGTT TCTAAACAT GAAGGACCTC CTGCAGAGAA ACCCTTGGAA GAACTCTCTG	360
	CTTCTACTTC AGGTGTGCCA GGCCTTTCTA GTTTCAGTC TGACCCAGCT GGCTGTGTGA	420
15	GACCTCCAGC ACCCAATCTA GCTGGAGCTG TTGAATTCAA TGATGTGAAG ACCTTGCTCA	480
	GAGAATGGAT AACTACAATT TCAGATCCAA TGAAGAAGA CATTCTCCAA GTTGTGAAAT	540
20	ACTGTACTCA TCTAATAGAA GAAAAGATT TGGAAAACT GGATCTAGTT ATAAAATACA	600
	TGAAAAGCCT GATGCAGCAA TCGGTGGAAT CGGTTTGGAA TATGGCATTT GACTTTATTC	660
	TTGACAATGT CCAGGTGGTT TTACAACAAA CTTATGGAAG CACATTAAAA GTTACATAAA	720
25	TATTACCAGA GAGCTGATG CTCTCTGATA GCTGTGCCAT AAGTGCTTGT GAGGTATTTG	780
	CAAAGTGCAT GATAGTAATG CTCGGAGTTT TTATATTTT AAATTTCTTT TAAAGCAAGT	840
30	GTTTTGTACA TTTCTTTTCA AAAAGTGCCA AATTGTCTAG TATTGCATGT AAATAATTGT	900
	GTTAATTATT TTAGTGATG ATAGATTCTA TTTACAAAAT GTTTGTTTAT AAAGTTTTAT	960
	GGATTTTAC AGTGAAGTGT TTACAGTTGT TTAATAAGA ACTGTATGTA TATTTGGTAC	1020
35	RGGCTCCTTT TKGTGAAYCC TTA AAACTC AACTCTAGGA RGCAACTACT GTTTATTATA	1080
	CTAAARGCCT GAAAACCTC CAGGCCAGAC TGCTAAGCTC TGAAATTCCT GAGAGGTCTC	1140
40	AGACCGGGAT TCTACTTGT CCAAGAAAG GTAAAGCTTC TAAACCATCT TATCTTGTG	1200
	TCCAAGCATG AACACAGGAG CATGTYAAGA AAATCTTTAC TACTTTCTYC CATGCCGAGA	1260
	AATCTACATA TTTTGAATTA GAAACACCT CACACCCACT TGAAGATTTT TTTCTGGGA	1320
45	ACATTATGTC CCGTAGATCA GAGGTGGTGT TGTCTTTTGT CTTCTACTGG CCATTGAGAA	1380
	ACTTTGATGA TAAAAAGAA CGGTATAGAT TTTTCAAACG TATATAAAAT ATTTTATGT	1440
50	TATATGTTAT GCCATAACTT TAAATAAAA ATAGTTTAAA ATTCTATGCT AGTGGATATT	1500
	TGGAACTTTT TCTCAAACA AACACCCAC ACTGACTTCA GCAAAACCT AAAACTAGCT	1560
	ACAGATTACT ACTACGAATG AATCATYAAG TTTGTGTCT GCAACAATTT AGAAGCACTA	1620
55	AGCCCAATA TCAGGAAATG TGTGTATGAT GGAATTTTCT AGGACAAAAC AGATCAAGAT	1680
	TAAAACAGGA TCAAGGATTA ATGGTATAAA AATGCTCTAC TAAAACAGGA TCAAGGATTA	1740
60	AAACAGGATC AAGGATTAAT GGTATAAAAA TCTCTACTGG TTACCGGGTG GCMGGCCAT	1800

ACAGGCTAGT GGTGCATGGA TAGTTTAGTT TGGNACGGT AA

1942

5

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 770 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

GGCAGGAGCC CTATGCTGTT CTTGTGATAA TGAGTGAGTC TCACAAGATC TGGTGGTGT 60

ATAGGCATCT GGCATTTCCC CTGCTGACGC TCATTCTCTA TCCTGCCACC CTGGCAAGAA 120

20

GTGTCTTCTG TCATGATTGT AAGTTTCTTG AGGCCTCCCC AGCTATGTAG AACTGTGAGC 180

CAATTAACC TCTTTTCTCT ATAAATTATC CAGTCTTATA TATTTCTTCA TAGCAGTGTG 240

25

AGAACAGATA ATACCGTAAA TTGGTATCAC AGAGAGTGGG GTGTTGCTAT AAACACATCT 300

GAAAATGTTA AAGCAAATTT GGAAGTGGGT AACAGGCAAA GGCTGGAACA GTTKGAAGAA 360

CAGTTAAGAA GAAGACAGGA AAATATGAGA AATCTTGAAA CTTCTAGAG TCTTAAAGGT 420

30

CTCAGAAGAC ATGAAGATGT GGGAAGCTTT GGAAGTCTCT AGAGACTTGT TTGAATGGCT 480

TTGACCAAAA TGCTGATAGT GATATGGACA ATGAAGTCCA GGCTGAGCTT ATCCAGACAG 540

35

ACATAAGAAG CTCGCTGGGA ACTTGAGTAA AGATCACTCT TGCTAGGCAA AGAGACTGGT 600

GGCCTTTTTT CCTCTGCCCT AGAGATCTGT GGAAATCTGA ACCTGAGAGA GATGATTTAG 660

GGTATCTGGC AGAAGAAATA TCTAAGCCGC AAAACCTTCM AGAGGAAGCA GACCATAAAC 720

40

GTTTGAAAAA TTTGCAGCCT GACNATGGGA GACCAAAGTT AAACCCAATT 770

45

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 519 base pairs

50

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

55

GAATTCGGCA CGAGCTGAGA GGCACAGGAG CAACAGCCAG TGGCCCTTCC AGAGGACCAC 60

TGGGGTCACA GACTTCARAC CTGATGACCT GGGCTCAGAT CCCAGCTCTG CACCTACCAG 120

60

CCGTGTGACA AGGTGTCCTC TCTGAGCTC AGTCACACAC TGCCCTTACG GTTGGGCCTC 180

414

ATGGAGCTGT TTGTGAAGGT TAAATGGGAA GACATAAACC ACTTAGCCCA GAGCCAAGGA 240
 CATGCTGAAT ACCATAATGG TGGCCTCCTT TGGCGCTGTG CTGGTGCAGG TGTGCCGAGG 300
 5 AAYTGGGCAG GGGTGACAGA TACCTCTTCT AACCTAGTTC CTTTCCAAGA ACCTAATTGG 360
 TGTCTCTCCC TCCCCCAGGC AATGGAAGG AGGAGGCTGG GCCCCAGCCC CAGAATACGG 420
 GAGGTTTCTC ACCCTGGTAG GGAAATTGCT GGGTTGGGGG TGTGGGCAAC CACAGTCATC 480
 10 GTCTCTCTGC AGGACGGATG AGGCTTTGCT GACAGAGGC 519

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 753 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

GGCAGGAGCG GCACGAGCAG CCAGTTGCTG ACTGGCACAT GGCTCCAGC GTCCCCGGCTG 60
 GTGGGCACAC TAGAGCCCGA GGGATCTTCT TAATTGGTAA ATTGGATCTT GAAGCTTCAC 120
 30 TGTTTAAATC TTTTCAGTGG CTTCCTTTTG TACTTAGAAA AAAATGCAAC TTCTTCTGCT 180
 GGGACTCATC CGCTCACAGC CTTCCTCTCC ACCCTCTCTC TGCCTCATGC TCTGCCCTCG 240
 CCTGCCATGC CTCCGATACT CACCTTTTGT ACCCCAGCAC CCGTGCCCTC TGCCCTCGA 300
 35 TCTTTGCCCTG CCTGGTTGCT COTCACTCAG TGTTCAGGAC AAATGCTCCT GGCCCTACCC 360
 CATCTAGCCA GTCTAGCCCG GTCTTCCCTG TCTTCCCTGT TTCATTGATG GCTCTTATTG 420
 40 TTTGTTWACT TGTGTGCTGT TGACTTTTAA CTCTCTCACT CCCCAGTGA ATGCAAGCGA 480
 TCTCCCAAGC TCTAGAATT GTTCCTGCCT CTTCAAGGC CTTACGCTG TGTGTGCTCG 540
 TGCCGAATTC GGCAGGAGG TATGTGCACT TGCTGGTATG TATGTAGGTG TTTGCTAACA 600
 45 CATACGTGCA CACGCAGAAT GGTTCAGGG GACTGCACAG CCTCTAGTTC GCAGCCCCCA 660
 CCGCTCCCTT TGSCCCTGCA CTCTCCCTC TCTGAGCTGC ATTGGCATGA AAGGCTGCAN 720
 50 GGTTCCTGAN CCGGCAAGG NCACCTCTG GGA 753

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1400 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

415

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

5	GGCAGGTTT ATTAAACCT ATTATGAAA AGTCACTTTG GTTGGCATTG AAAATTACAT	60
	CATCTTTAAA GCAGTATTTG TCCCAGATG GACTCATGAC TAGCAAAGAC TAGGTTGATT	120
	GGAAGGCATA GGGTGAGAGA ATGGGAGAT GRAGTGGAGG CGGGTTGTTA AAGTGCTGTC	180
10	AGTGAGTGAT TTGTGCTACT TGAACAATGG TCCATGTTTG GGGGCATATT GTGTTTCATA	240
	AGAAGTGAAA GGTATTTGCA AAGTAGCTA CAAATGACCC ATAAATCTGT TAACAACAGT	300
15	CCTTAATATG CAAAGATGAA AAACAAGCAT TACTGCTACC CAAAGGGAAC TGCTGCTTGG	360
	TGATGTGCAG ATGGGGCTGT TGGTTAAGAG AGCTATTACA GTTTTCTCTT CTAGGTTTC	420
	ATAGGAGCTA GTTACTGACA TGAGATTGTT TTATCTTTT CAATACAGAT CTCTTGCTTT	480
20	GAGTTAGTTC TGAGGATGGG AGTAATAAAG GAGTTTTTGG TTTTTTGTG TGTGTTTGG	540
	TTTTGGCTCC TTAGTAATAC TCGTCTGCA TTTATTTCTA TTATTTCTCA AAGAAAGGAA	600
25	ACCAACTGAA ATGTTTGCTT TAACAACAT TTAAATAAGT TCTCTGGGTT TTTTTTCCC	660
	CTTTTAAAAA AATTAGCATA TACCATGCA ATAAAGCAAC TAATGTTAAC TATTGTATGC	720
	TACAACTTAA GTGATTTTTC TAAGCAAGCA CATGTCATT GRAAGTATTA TTGAAAAGCA	780
30	TCATAGTCAC ATTGAATTG TGAAGGCCAA AGAAATTGAA GGGAGTGATA TTTCATTTT	840
	ATGATATTCA CATATTTAGT AAATTTTGTG TACAAGAATA CCAGGCAGAG TGTTTTACCC	900
35	ATGGAAACAG GTTTCAGATT ACTTTGTTTT TACTGTTAGA GTCTCAAGTT TAGAAATGCT	960
	AACACTTAAA TCAGTTTTTT TCTCACTATA CTTGAAGATT GTTAATATTT TGATATCTTC	1020
	CTAGCTTGAT GGAATTTTAA CATATCTTCA GATCTGTGAC AGTGACAGCC AATAGGACTG	1080
40	ATAATATTAG CTTCAAGCA ATAATATCCA GGGTTAAAT AAAATCATA GTGAAAGTAC	1140
	GATTGTAAAA TTATGCTATA TTAACTTTTA AGTCTGTAAT AACTTGACAT CAAATGTGA	1200
45	TGTAATTACC ATAAATAATG GCTAGCGAGA ACATCTTTGG AAATTCCTCA ATTACCTTTC	1260
	TTACTACACT GTTTGCAGAA TGAATGTAGA AATGATCCTG TTAGCTTTCT GAATGTTCTG	1320
	TGGTTGAATG TGTGTTTGGT TAAATAAGC TTTTGGTATT TGTGTTAAATW AAAAAA	1380
50	AAAAAAAAA AAAAACTCGA	1400

55

(2) INFORMATION FOR SEQ ID NO: 165:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2153 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

5	CAGGCCTCAG GGCCTCTGGT GGCTCTGGCC CAGACACTAT TTGCAGTTCT TGTGCTATGG	60
	GTGGGAGTCT TCTTCTCAA GTTCGGCCAG CTGTGCTGTG NCTGGATGGG CTGCTCCTCC	120
10	CAGGGCTCAA GGGCTGTGGT CCGCTCAGGG TCTCATTTCC CCAGGCCAAG TTCAAGGCAG	180
	CAGCCCTTTG TGAGGCGGTC TTGGCCCTGG GCTGGAGGGA GAACTTTAAG CTTTTTTGCT	240
15	CACAGGGAGC TGCTATGGGC CCTGGGTGCA GGTGCCACACA TTCTGCTAAT GAGAGCTTTG	300
	TCTGATCAGT CCTGGGTCCA TCAGTTTGTG CATGTGTCCG GCTGCCAGGC CGTCCCTTGG	360
	GATCCTTCCC CTGGGGTGTA GCCTTGTTCA TTAGTATATA CTCATTCTTT CATGCTTTTC	420
20	TCAGCAGAAC ACTTCCACTT CTGAGGTGAG CTTTTCGCCC RTGCCCTTCC TCCACAGGTG	480
	TTGCCCTTTT ATAAAGACCT GATAGCAGAA TAAATTGGTG TTTCCCTGTT GACCCAGCAC	540
25	CATTTCTGTG GGCCTAGAAT ATGGCCCTCA ACCCTTAGAG TGGGGCAGTG AGGGCTTGAG	600
	GAGTGACCCCT TCCTTTCTCA TGGTTTTAGT CATTTTGGCT GCCAGCCCTT AATGGCACAG	660
	ATCTGCTGCT TCTAACAGAT GGGCAGGAGG TGACACCGAT TTCAGCCATT GCCAAGGTTA	720
30	GCACCCCTCT CTTTGAGCCT AGGGCCACAC TGTTCAATTGT CACTTTAGGC AAGTGCCTGT	780
	TTGGCTTTAA AGGTAAGCCT GCCAGCTGTG AGAAGCCCTG GTAAGTATG GACTCATTTT	840
35	CTGGTCCCTA AAGATGCAGC CTCTTAAGGG CTCCTTGATG GATGCCATCT CTCCTAGCCC	900
	CCAGCCCTGG TGCCACTGGT GGGCAGGTC CCATTCTTTG GGGCTGGGAG GGACAGCTTG	960
	CCTGTTTCTG GTCACAAATT ACAGTCTTCT CTCCTGTACC ATTCTGTGGC TTCAGCATGG	1020
40	GGCAGTAGC CTTTCATTAG TGTAGATAGT CATTCCTGG TAGGGTGGAG GGTAAGACAT	1080
	AGGGTCTGGA ACTGTTTGGG ACCTTTTGGG GATGTCTGT GCTTCCAGG TTCCTMGATT	1140
45	CTGGGAGGAG AGGCTGCCGC ATTCTGTGTC TCCTCACAGC GAGCAAAGCT GCACCCACTT	1200
	ACATTCACTA TTTTCTGGC ACTACAAAGA GTGGGAAGGC CTGGGATTTG CTGCTGCTCC	1260
	CTTAGAGCAG GGGCCCTTCT TTCAGCACTT TGGACACCTG GAGACCCAGC CCTGTTATTT	1320
50	AATGGTAGTG GGCAAGTGTG TGTGCATACT GTCTGCCACT GCTTTCTCCC TGGCCCATGC	1380
	CAGAGAGCCC TGTCCCTGCC AGGCCAGCC TTCTTAGCCC CAACTTGGGA ACAAAGTGCA	1440
55	ACATGGGATC ATGGGTTGGG GTGCTCAGGT GAGCCCTCTC TATAGTGCTT CCTGGGGCCA	1500
	AGCTGACACC AGCCCTGAG GGTGGGGTGG GACGGGTGGT GCTTAAAGA GGAAGGGGAC	1560
	CAGTGTAGCA ACTTGGCAGG GACCCACCCC CTCCTCTCT GGGCTGTGC AGTGAGCATG	1620
60	GGGATTCCCA TCAAGGGGCC TGGCACCTGT GCTAGTTACG TAGCCGCTGN TCACGGGCTC	1680

ACTCCTGACC ACATGCACGT TCCCTAGATG CAGACTGCTT TGAACCTTAA AGCTGTACAA 1740
 TTTGGTTATG TTTGTGCTGA CTTAAAATAT ATTTTAATGA CGAAAAAATA ATGGAGAACC 1800
 5 CTGGGAAGGA CCTGTTCTT TTGCTTCTCG GGGAACTGTA AGCCCTCGCG TTCTGGGAAT 1860
 CGCTCTCTGC TGCTCTTTCC TGAAGCTAA GCTGTCTCC ACCGCCCGAG GCCTGCGCCG 1920
 10 GTGCTCCCGC CGCAGTTCCG TTTGCTTTGG ACCTTGCGTG CGGGGGAGGG GGTGCTCGGT 1980
 CCGAGCCCGC TCCPTTCTGT ACACCTAGCG CTGCCCCGCC CGTTTGTGTC TGAGGTGCTG 2040
 TATGTCAAAA ATAAAGCCGC TAGAACCGGA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2100
 15 AAACTCGAGG GGGGGCCCGT ACCCAATTAA CCGMTATGA TCTATAAGC GTC 2153

(2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1251 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

30 GCCCACGCGT CCGCCCACGC GTCCGGCGGT GCGGAGTATG GGGCGCTGAT GGCATGGAG 60
 GGCTACTGGC GCTTCCTGGC GCTGCTGGGG TCGCACTGC TCGTCGGCTT CCTGTCCGTG 120
 35 ATCTTCGCCC TCGTCTGGGT CCTCCACTAC CGAGAGGGGC TTGGCTGGGA TGGAGCGCA 180
 CTAGAGTTTA ACTGGCACCC AGTGCTCATG GTCACCGGCT TCGTCTTCAT CCAGGGCATC 240
 GCCATCATCG TCTACAGACT GCCGTGGACC TGGAAATGCA GCAAGCTCCT GATGAAATCC 300
 40 ATCCATGCAG GGTAAATGC AGTTGCTGCC ATTCTTCCAA TTATCTGTGT GGTGGCCGTG 360
 TTTGAGAACC ACAATGTTAA CAATATAGCC AATATGTAGA GTCTGCACAG CTGGGTGGGA 420
 45 CTGATAGCTG TCATATGCTA TTTGTTACAG CTTCTTTTCA GTTTTTCACT CTTTCTGCTT 480
 CCATGGGCTC CGCTTTCTCT CCGAGCATTT CTCATGCCCA TACATGTTTA TTCTGGAATT 540
 GTCATCTTTG GAACAGTGAT TGCAACAGCA CTTATGGGAT TGACAGAGAA ACTGATTTTT 600
 50 TCCCTGAGAG ATCCTGCATA CAGTACATTC CCGCCAGAAG GTGTTTTCTG AAATACGCTT 660
 GGCCTTCTGA TCCTGGTGTT CGGGGCCCTC ATTTTTTGA TAGTCACCG ACCGCAATGG 720
 55 AAACGTCTTA AGGAGCCAAA TTCTACCATT CTTTCATCCAA ATGGAGGCGC TGAACAGGGA 780
 GCAAGAGGTT CCATGCCAGC CTA CTCTGCGC AACACATGS ACAAAACAGA TTCAGAGTTA 840
 AACAGTGAAG TAGCAGCAAG GAAAAGAAAC TTAGCTCTGG ATGAGGCTCG GCAGAGATCT 900
 60

ACCATGTAAA ATGTTGTAGA GATAGAGCCA TATAACGTCA CGTTTCAAAA CTAGCTCTAC 960
 AGTTTTCCTT CTCCTATTAG CCATATGATA ATTGGGCTAT GTAGTATCAA TATTTACTTT 1020
 5 AATCACAAG GATGCTTCT TGAATAATT TGTATTGATT GAGGCCTATG AACTGACCTG 1080
 AATTGGAAAG GATGTGATTA ATATAAATA TAGCAGATAT AAATGTGGT TATGTTACCT 1140
 TTATCTTGTT GAGGACCACA ACATTAGCAC GGTGCCTTGT GCAKAATAGA TACTCAATAT 1200
 10 GTGAATATGT GTCTACTAGT AGTTAATTGG ATAAACTGGC AGCATCCCTG A 1251

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 882 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

GACSMCTTAG AACTATGGTC CCCCCGGACT GCAGGAATTC GGCACAGCGG CTGCGGGCGC 60
 GAGGTGAGGG GCGCGAGGTT CCCAGCAGGA TGCCCCGGCT CTGCAGGAAG CTGAAGTCAG 120
 30 AGGCCCGGAG AGGCCCCAGC CCGCCCCGGG CAGGATGACC AAGGCCCGGC TGTTCGGCT 180
 GTGGCTGGTG CTGGGGTCGG TGTTCATGAT CCTGCTGATC ATCGTGTACT GGCACAGCGC 240
 AGGCCCGCGG CACTTCTACT TGCACCGTC CTTCTCTAGS CCGCACACGG GCGCGCGGCT 300
 35 GCCCACGCCC GCGCCGGACA GGCACAGGGA GCTCACGGCC GAYTCCGATG TCGACGAKTT 360
 TCTGGACAAK TTCTCAGTG CTGGCGTGAA GCAGAGTGAC YTTCCAGAA AGGAGACGGA 420
 40 GCAGCCCCCT GCGCCGGGGA GCATGGAGGA GAGCGTGAGA RGCTACGACT GGTCCCCGGC 480
 CGAMGCCCGG CGCACCCAGA CCAGGCGCGG CAGCARGCGG ANCGGAGGAR CGTGCTGCGG 540
 GGCTTCTGGG CCAAYTCCAG CTTGGCCTTC CCCACCAAGG AGCGGCGATT CRACGACATC 600
 45 CCCAACTCGG AGCTGAGCCA CTTGATCGTG GACGACCGGC ACGGGGCCAT CTACTGCTAC 660
 GTGCCCAAGG TGGCCTGCAC CAACTGGAAG CCGGTRATGA TCGTGCTGAG CGGAAGCTGT 720
 50 GCACCGCGTG CGCTACCGC GACCCGYTGC GATCCCGCGC GAGCACGTGC ACAACGCCAG 780
 CGCGCACTGA CTTCAACAAT TCTGGCGCGG CTACGGGAAG TCTCCCCAC CTCATGAAGT 840
 55 CAAGCTCAAG AATACACCAA TTCTTCTGC GCGACCTTC TG 882

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1208 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

GGGAAACTCA AAAGCATGAT GGAATGGTTG ATGGAGCCAG AGCCTAGAAG TRAAGGGATA 60
 CAGAGTGAAG ATAGAGGTAT TTACGTATAT TTAAATATTA GCTTTGGAAT TACGTAGGGA 120
 TTCTTAAGAA AAGATCATGA CAGGACAGCC ACATTTGGTA AAATGTCAGG GCAGCCAGTG 180
 CATGGTCCTC CTGGGGCTCC TCAGTTGACG GGTTTAAATC ATTTCTGAT CCCCCTGCCC 240
 TGGTTTGAGG AATGCATACA GTACGTGAAA TGGCTGTGCT ATGAGTTCCA ATGGGCAATC 300
 AACCTGGGTA AATCCAGAT TAATGATTAG TTCTAAAGAT CCAGTTGAAG TTCTAGAGTC 360
 GGAATTTTCC GTCAAGCARG TCAGCACAGC TTTATGCCCTG TTCCTCTAAT AAGGATAGGT 420
 AACAAATAGC TGTGKTWCA CAGCTAGGAR GATAACCAA TCTAGAGTTC TTGARTCTCA 480
 TTTAATAAAT AATATTATG AGTACCAACT GCATATTTCA GGCAGTGCAT TTGACTCTGT 540
 TAAATACTGA TYCCTAKGA CMSCCAGWTC AGAWAACMTT AATCTGTCTG ATCAATAAAC 600
 AGCTTGACTT AGAGRGGTAA AATAGCTTGC CACAGGTWAC CCAATTAGTA GGTAAACAGG 660
 ACAGAATAAC AGTGCAGTTA AAATCTTAGA CTGGAGACTA ATTGCATAAG TTGAATTTC 720
 AGTTCTGCTA TGTAAATTTG GGTGAGTACC TTAATTYACC TGAGTCTCGG TCTTTATATC 780
 TGTAGAATGG AGCTAATGAT ATTACTTAAT TTGCTTTATG TGAGATTAAA TGTACTAATA 840
 TATGTAAATC ACTTACAACA GCATTTGACA TATTTGACAT ACTTAATATA TTTGCTACTA 900
 ATACTATTAG CAACAGCATT CTGATTTTCC AAGTTGAAAT TCACTGTTTT CTTTMTTACT 960
 TTGCCATAAT TTACAATGTT GTGCTCTGTA AACCATAAAT TTCCCTGAGG TGTGTCAGG 1020
 TTAACAAAAA ATCACTATGG CCCCCANMA CTTGGAAAAT AGAAATGAGA CCAGCTTCAT 1080
 CTATATTCTT TACTGCAAT AACTTAGAAT TGTAATAGGC TAATATGTAC TGGGACTTCC 1140
 AATTGCGGAA TATGACAAAA ATAATACTAT TTAGCTAAAA CATATACAGA ACTTATTTTT 1200
 CCTCTGAA 1208

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1307 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

5 GGCACGAGAG AAAAGAGGTT GACAATGTTT TCTAGCAGGC AGAATGTGCA TACATGTTTT 60
 CATGARTGTC CTTTGGGTGC TGTTCCTTTT AAATCCTCTG TGCACAGGGC TCTGGCCTTT 120
 ARTAACTGT TTTTCTGTCT TACGTCATGC TGACTGGGTG CTAGGGGCTG ATTACAAAGG 180
 10 GGAAGAGTTG AACAGACATC AGGGGCCGAT GAAACCAAG GACTAGGAGT CAGGAGAACA 240
 AGTCAGGGAT TAGGAGACAG CGGTTTGTTT TATGTATC CAGCTGGAGG ACTCCTAGGG 300
 GCACGACGAG GAGGAATACC AGGCCACCG AGGGGCAGGA GTCTCACAGT GGAGGGCACA 360
 15 CTCTAACAGA TGCCAGCTGA ACCCTCGCTG GCCTGGATG TCATACGAGT TGGGGACCAG 420
 AAATCTGGGC TCAGAGAACC CGTCCAGGGA GATTTGAACC CATGGGTTAT CTCTAGAGT 480
 20 TGATACTGAT AATATATTTT AATTTTATT GATGTTTAAT ACCTTCTGAA ACAGGAGGGT 540
 AAGATCAGAT GGAAGCCCT TGTGTTGAAG GATCTTGGGA ACCTTGGTGG TTTTTTTTTT 600
 TTGGTTTTTT TTTTTTGAT CGAGCTGTGG ACATCCTTCT TAATTCGATT NTCAGGATTT 660
 25 GTTTAACTAA AAAGTTCCCA AACACAGAAA GGGCCTCCCC ACCTGCTTTG GGGAGCTGTC 720
 TGTCTGGGA GTGCCAGGCA TCCSATGGGA CCCATCACTG CCAGTGTCTG TGCCTCCCAG 780
 30 AGGTCAGCCC TGTGTCTGCC CTGGCTCTGT CTCCTCTGTG ACAGGGCAGA GCATTTCTGG 840
 TCAGTTTCTC CATGGTGCCT CCCACCCCTT TGTAAAGTGG ATGGACATGA TGGAAATCAG 900
 TTGTCTCACC CTGATAGCCT GGGTCTTGT ATTCACTTTA CCCGCACTCA GACACAGGGG 960
 35 ACCTTGAAGC AGTTCTCGGT GTGTAGAGTC CACGTGACAG TCCCCACAGC CTCCCCAGAT 1020
 AGCTGTGTGC CTGTGCGCTA CTGCTGTGCC ATTTTCCCAA CTTNGGCGTT TCACTAAATG 1080
 40 CAGCTGATCT CTCTCTCTGT GCACTCGTGA TCCATGTTGA ACAATACATG TAGGTTCTTT 1140
 TTCCACCCAA TGTAAAGACA TGATATACTG TACGTGGAA AGCATTTACC TTATTTATAT 1200
 45 ACCTGAATGT TCCTACTACA CAAATAACA TATATTAAAT WCTAAAAAAA AAAAAAAAAA 1260
 CTGGAGGGGG GGCCCGGTAC CCAAATCGCC GGATAGTGAT CGTAAAC 1307

50

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 1624 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

	GGCACGAGGT CCGCGCCGCG GCGCCCTGGA ATTGTGGGAG TTGTGCTGTC CACTCGGCTG	60
	CCGGAGGGGA AGGTCCCTGA CTATGCGCTCC CCAGAGCCTG CTTTCATCTA GGATGGCTCC	120
5	TCTGGGCATG CTGCTTGGGC TCCTGATGGC CGCCTGCTTC ACCTTCTGCC TCAGTCTCA	180
	GAACCTGAAG GAGTTTGCCC TGACCAACCC AGAGAAGAGC AGCACCAAG AAACRGAGAG	240
10	AAAAGAAACC AAAGCCGAGG AGGAGCTGGA TGCCGAAGTC CTGGAGGTGT TCCACCCGAC	300
	GCATGAGTGG CAGGCCCTTC AGCCAGGCCA GGCTGTCCCT GCAGGATCCC ACGTACGGCT	360
	GAATCTTCAG ACTGGGGAAG GAGAGGCAAA ACTCCAATAT GAGGACAAGT TCCGAATTA	420
15	TTTGAAAGGC AAAAGGCTGG ATATCAACAC CAACACCTAC ACATCTCAGG ATCTCAAGAG	480
	TGCACTGGCA AAATTCAAGG AGGGGGCAGA GATGGAGAGT TCAAAAGGAG ACAAGGCAAG	540
20	GCAGGCTGAG GTAAAGCGGC TCTTCCGCCC CATGAGGAA CTGAAGAAAG ACTTTGATGA	600
	GCTGAATGTT GTCAATGAGA CTGACATGCA GATCATGGTA CGGCTGATCA ACAAGTTCAA	660
	TAGTTCCAGC TCCAGTTTGG AAGAGAAGAT TGCTGCGCTC TTTGATCTTG AATATTATGT	720
25	CCATCAGATG GACAAATGCG AGGACCTGCT TTCCTTTGGT GGTCTTCAAG TGCTGATCAA	780
	TGGGCTGAAC AGCAGAGAGC CCGTCGTGAA GGAGTATGCT GCGTTTGTGC TGGGCGCTGC	840
30	CTTTTCAGC AACCCCAAGG TCCAGGTGGA GGCCATCGAA GCGGGAGCCC TGCAGAACT	900
	GCTGGTCAIC CTGGCCACGG AGCAGCCGCT CACTGCAAAG AAGAAGGTCC TGTTTGCACT	960
	GTGCTCCCTG CTGCGCCACT TCCCCTATGC CCAGCGGCAG TTCCTGAAGC TCGGGGGGCT	1020
35	GCAGGTCTTG AGGACCCTGG TGCAGGAGAA GGGCACGGAG GTGCTCGCCG TCCCGCTGGT	1080
	CACACTGCTC TACGACCTGG TCACGGAGAA GATGTTGCGC GAGGAGGAGG CTGAGCTGAC	1140
40	CCAGGAGATG TCCCCAGAGA AGCTGCAGCA GTATCGCCAG GTACACCTCC TGCCAGGCCT	1200
	GTGGGAACAG GGCTGGTGCG AGATCACGGC CCACCTCTTG GCGCTGCCCC AGCATGATGC	1260
	CCGTGAGAAG GTGCTGCAGA CACTGGGCGT CCTCCTGACC ACCTGCGGGG ACCGCTACCG	1320
45	TCAGGACCCC CAGCTCGGCA GGACACTGGC CAGCCTGCAG GCTGAGTACC AGGTGCTGGC	1380
	CAGCCTGGAG CTGCAGGATG GTGAGGACGA GGGCTACTTC CAGGAGCTGC TGGGCTCTGT	1440
50	CAACAGCTTG CTGAAGGAGC TGAGATGAGG CCCCACACCA GGAAGTGAAT GGGATGCCCC	1500
	TAGTGAGGCT GAGGGGTGCC AGCGTGGGTG GGCTTCTCAG GCAGGAGGAC ATCTTGCCAG	1560
	TGCTGGCTTG GCCATTAAAT GGAACCTGA AGGCCAAAAA AAAAAAAAAA AAAAAAAAAA	1620
55	AAAA	1624

422

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2003 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

10	GGCACCAGCC AGCTTGCAGG AGGAATCGGT GAGGTCCTGT CCTGAGGCTG CTGTCCGGGG	60
	CCGGTGGCTG CCTCAAGGT CCCTCCCTA GCTGCTGCGG TTGCCATTGC TTCTTGCTTG	120
15	TTCTGGCATC AGGCACCTGG ATTGAGTTGC ACAGCTTTGC TTTATCCGGG CTTGTGTGCA	180
	GGGCCCCGGCT GGGCTCCCCA TCTGCACATC CTGAGGACAG AAAAAGCTGG GTCTTGCTGT	240
	GCCCTCCCGAG GCTTAGTGT CCCTCCCTCA AAGACTGACA GCCATCGTTC TGCACGGGGC	300
20	TTTCTGCATG TGACGCCAGC TAAGCATAGT AAGAAGTCCA GCCTAGGAAG GGAAGGATTT	360
	TGGAGGTAGG TGGCTTTGGT GACACACTCA CTTCTTTCTC AGCCTCCAGG ACACTATGGC	420
25	CTGTTTTAAG AGACATCTTA TTTTCTAAA GGTGAATTCT CAGATGATAG GTGAACCTGA	480
	GTTGCAGATA TACCAACTTC TGCTTGTAAT TCTTAAATGA CAAAGATTAC CTAGCTAAGA	540
	AACCTCCTAG GGAAGTAGGG AACCTATGTG TTCCCTCACT GTGGTTTCCT GAAGCCAGTG	600
30	ATATGGGGGT TAGGATAGGA AGAACTTTCT CGGTAATGAT AAGGAGAATC TCTTGTCTCC	660
	TCCCACCTGT GTTGTAAGA TAACTGAGG ATATACAGGC ACATTATGTA AACATACACA	720
35	CGCAATGAAA CCGAAGCTTG GCGGCCTGGG CGTGGTCTTG CAAATGCTT CCAAAGCCAC	780
	CTTAGCCTGT TCTATTGAGG GGCAACCCCA AAGCACCTGT TAAGACTCCT GACCCCAAG	840
	TGGCATGCAG CCCCCATGCC CACCGGGACC TGGTCAGCAC AGATCTTGAT GACTTCCCTT	900
40	TCTAGGGCAG ACTGGGAGGG TATCCAGGAA TCGGCCCTG CCCCACGGGC GTTTTCATGC	960
	TGTACAGTGA CCTAAAGTGG GTAAGATGTC ATAATGGACC AGTCCATGTG ATTTCACTAT	1020
45	ATACAACCTCC ACCAGACCCC TCCAACCCAT ATAACACCCC ACCCCTGTTT GCTTCTGTGA	1080
	TGGTGATATC ATATGTAACA TTTACTCCTG TTTCTGCTGA TTGTTTTTTT AATGTTTTGG	1140
	TTTGTTTTTG ACATCAGCTG TAATCATTCC TGTGCTGTGT TTTTATTAC CCTTGGTAGG	1200
50	TATTAGACTT GCATTTTTT AAAAAAGGT TTCTGCATCG TGAAGCAAT TGACCCAGAG	1260
	TGGAACCGGT GGCCTATGCA GGTGGATTCC TTCAGGTCTT TCCTTTGGTT CTTTGAGCAT	1320
55	CTTTGCTTTC ATTCGTCTCC CGTCTTTGGT TCTCCAGTTC AAATTATTGC AAAGTAAAGG	1380
	ATCTTTGAGT AGGTTCCGTC TGAAAGGTGT GGCCTTTATA TTTGATCCAC ACACGTTGGT	1440
	CTTTTAACCG TGCTGAGCAG AAAACAAAAC AGGTTAAGAA GAGCCGGGTG GCAGCTGACA	1500
60	GAGGAAGCCG CTCAAATACC TTCACAATAA ATAGTGGCAA TATATATATA GTTTAAGAAG	1560

423

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60

GGTCTCCATT TGGCATCGTT TAATTTATAT GTTATGTTCT AAGCACAGCT CTCTTCTCCT 1620
 ATTTTCATCC TGCAAGCAAC TCAAAATATT TAAAATAAAG TTTACATTGT AGTTATTTTC 1680
 AAATCTTTGC TTGATRAAGTA TTAAGAAATA TTGGACTTGC TGCCGTRAATT TAAAGCTCTG 1740
 TTGATTTTGT TTCCGTTTGG ATTTTGTGGG GAGGGGAGCA CTGTGTTTAT GCTGGAATAT 1800
 GAAGTCTGAG ACCTTCCGGT GCTGGGAACA CACAAGAGTT GTTGAAAGTT GACAAGCAGA 1860
 CTGCGCATGT CTCTGATGCT TTGTATCATT CTTGAGCAAT CGCTCGGTCC GTGGACAATA 1920
 AACAGTATTA TCAAGAGAA AAAAAAAAAA AAAAAACTCG NGGGGGGGCC CGGTACCCAA 1980
 TTGCCCCAT ACTGAGCCNA TTC 2003

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 786 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

30
 35
 40
 45
 50
 55
 60

GGCACAGCGG CACGAGAAGA CTTTGGTGT TAAGAGATTA ATGTGTTAGC CAGAACAAC 60
 CATTTCTCTA CCGTGTGTGA GTCCATTTAT CTTTAAAGAT TTTCTATTGG AATAATTTTG 120
 AAATTACTTT CTTAGTTTTT TCATTAAAA ACTAAGAAAA TGCTTTGTTT ATTATGAATT 180
 GCTATTTCTC TTGATTATTA TTCTTGGAGA AAGTCTATCA GACGTAATTC TTCTGATTTG 240
 CTCTAGGCT AGAGGAAAT GTGAAAGATG ACAAATGAAA ATTTCAAAGG TTGTCAGTAG 300
 TATGACTTCT TTTATCGTTT GTCATTATCA CAAATATATC AACATAGGAC TTTTAAAAGA 360
 TATTTTGTAC ATATTGGGCC TTAGTAGGAT TTGTCATGAA TTTTTTTTTT CTTTATGCC 420
 CAGAGAGAAA GAGCAAAGAA ATAACCAAGG GTGATGTACT CGTATTGAAG GTTTACCAAA 480
 TAAGGACTGC TTTTATTATG AACTATAGTC TATATTCTAA GTAAATCAAT TTTTCTATTA 540
 TGTGTTTTTT GTTCCTGCAG GCAAGATCTC TGAACTTTAT GCAGAGGGTT CTTTTAAAAA 600
 AACAAAGTTG AATTTTTTTA TTTCTTGGAA TATTTTTTTT CATTGATTTT TCCCAAGTAG 660
 AGCAGATTCA AATCTCCTTT GTACCCATAG TCTTTTTTCT TTGCTATTA GCTCAGTATT 720
 CCGTTTCTAC ATTTTCCTTT CCTAGAACCA GTCAATAAAT GACAAAAAA AAAAAAAAAA 780
 ACTCGA 786

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1758 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

5	GGGACGAGCC CTGCCCACCT CTTGCAGCCT CTTGCGCCCC GCGGAGCTGG CGGATGGAGC	60
10	TGCGCAGCGG GAGCGTGGGC AGCCAGGCGG TGGCGCGGAG GATGGATGGG GACAGCCGAG	120
15	ATGGCGCGCG CGGCAAGGAC GCCACCGCGT CCGAGGACTA CGAGAACCTG CCGACTAGCG	180
20	CCTCCGTGTC CACCCACATG ACAGCAGGAG CGATGCGCGG GATCCTGGAG CACTCGGTCA	240
25	TGTACCGCGT GGACTCGGTG AAGACACGAA TGCAGAGTTT GAGTCCAGAT CCCAAAGCCC	300
30	AGTACACAAG TATCTACCGA CCCCTCAAGA AAATCATCGG GACCGAAGCT TCTGGAGCCC	360
35	CTTGGCGAGGC GTCAACGTCA TGATCATGGG TGCAGGGCCG GCCCATGCCA TGTATTTTGC	420
40	CTGCTATGAA AACATGAAAA GGACTTTAAA TGACGTTTTT CACCACCAAG GAAACAGCCA	480
45	CCTAGCCAAC GGTATTTTGA AAGCGTTTGT CTGGAGTTAG AAAGTTCTCT TCTTCAACAC	540
50	GTCCCTCCCC AGGGTGTTC TCCCTGTGAC CCAGCGCGCT CGACTTCGGC CCGCTTGCTC	600
55	ACGAATAAAG AACTCAGAGT TGTGTGTGCA ATGCACACCC AGACACACGC ACCCACACAC	660
60	ACGCGCGCGC ACACACATGC TTTTCTCTGT TCCCTCCGC TTTCTGAAGC CTGGGGAGAA	720
65	ATCAGTGACA GAGGTGTTTT GGTTTTATG TTAGTGGGT TTTCTTTGT ATTTTTTTTG	780
70	TTTGTTTTGT TTTTAAACAT TCAAAAGCAA TTAATGATCA GACATAGGAG AAACCTGAA	840
75	TAGAAACAAA ACTTTTGAAT GCTGGATTCA AAAAAAAAAA AAAGTTATCT GGACAGCTTC	900
80	TTTGAGACTA TTTAAAACT GGTACAACAG GTCTCTACAA CGCCAAGATC TAACTAAGCT	960
85	TTAAAAGGTC AAGAAGTTTT ATGGCTGACA AAGGACTCGC GCAACGCAGA AGGCCTTTCC	1020
90	CACCTTAAGC TTCCGGGGAT CTGGGAATTT TACCCCATTT CTCTCTGTT TGTCTGAGTC	1080
95	TCATCTCTCT GCAAGCAAGG GCTGAAATCA TTTTGTTTGG TTGTTTTGAG GGAGAGAGGC	1140
100	GGGGTGGGGG GGTGCAAATC TGCCAGCAGC TCTTACGTAA GGCATGTTTT ATGGGGAGG	1200
105	GCTGAGCTTT TATTTTCTCC TCTCCAGTGG GGTGGGCTTT TATTGTTTCT TGTTTGGGTT	1260
110	TGGAATGGAA ATATGGATAG CAGCATAAAG TACTTTTATT TTGACAAAAT TCATTTTTTT	1320
115	CAACAATGGA GACATAGATT TGACCCACAA TRACTTCTCC CCTCTCTTT TACTCTGCT	1380
120	CAAAAAGCAT CTCTCTCTCC ATTACCCAAC CTGGTCATA AGTGTGGCTG GCTGGTTTGC	1440
125	AGATATTTGT TCTGCTTTGT AAAAATGGC CATTAGTGCA TTTATTGAGA TGATCTCTAA	1500

425

AGAGCTATGC CCTGACCTAC CCTGATTCT ATGACATTGG GGGCCTTCTT TTGCTGAAAC 1560
TGCCTTACGT AATGGTTTTA CTCCTTGAAA GAGATTGAC GGAATCCATT TTATGCCAAG 1620
TGCTGCCCTG CACTGTTTCT GCAATATGTG GTGTATGCTG TGGTGATCTT GCTGGGAATG 1680
ATTATAAGTG TGTGTGTGGT GGGGGAGTGG GTATTACATG CATTGCTGAA GAGTCAAAAA 1740
AAAAAAAAA AAACTCGA 1758

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 888 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

CTGTTAGAAT GCCCAGTTTA COTGGATGGC AACCCACAG TGCTGCTGCC CACCTGCCCC 60
TCAATCCTCC TAGAATTCAG CCCCCAATTG CCCAGTTACC AATAAAAACT TGTACACCAG 120
CCCCAGGGAC AGTCTCAAAT GCAAATCCAC AGAGTGASMC ACCACCTCGG GTAGAATTTC 180
ATGACAACAA TCCCTTTAGT GAAAGTTTTC AAGAACGGGA AGSTAAGGAA CGTTTACGAG 240
AACAGCAAGA GAGACAACGG ATCCAACCTCA TGCAGGAGGT AGATAGACAA AGAGCTTTGC 300
AGCAGAGGAT GGAAATGGAG CAGCATGGTA TGGTGGGCTC TGAGATRAAGT AGTAGTAGGA 360
CATCTGTGTC CCAGATTCCC TTCTACAGTT CCGACTTACC TTGTGATTTT ATGCAACCTC 420
TAGGACCCCT TCAGCAGTCT CCACAACACC AACAGCAAAT GGGGCAGGTT TTACAGCAGC 480
AGAATATACA ACAAGGATCA ATTAATTCAC COTCCACCCA AACTTTTCATG CAGACTAATG 540
AGCGAGGCAG GTAGGCCCTC CTTGATTTGT TCTGATTCA CCATCAATCC CTGTTGGAAG 600
CCCAAATTTT TCTTCTGTGA AGCAGGGACA TGGAAATCTT TGTGGGACCA GCTTCCAGCA 660
GTCCCCAGTG AGCCCTTCTT TTACACCTGC TTTACCAGCA GCACCTCCAG TAGCTAATAG 720
CAGTCTCCCA TGTGGCCAAG ATTCTACTAT AACCCATGGA CACAGTTATC CGGGATCAAC 780
CCAATCGCTC ATTCAGTTGT ATTCTGATAT AATCCAGAG GAAAAAGGCA AAAAAAARA 840
AATAAARA ARAAAGGAGA TGATGATGCA GAATTCACCC AAGGCTCC 888

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:

426

(A) LENGTH: 2379 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

	GGCAGAGCTA GTGTGGACTC CATCCCCCTG GAGTGGGATC ACGNCTATGA COTCAGTCGG	60
10	GACCTGGAGT CTGCAATGTC CAGAGCTCTG CCGTCTGAGG ATGAAGAAGG TCAGGATGAC	120
	AAAGATTCTT ACCTCCGGGG AGCTGTTGSC TTATCAGGGG ACCACAGTGC COTAGACTCA	180
15	CAGATCCGAC AACTGGGCAA AGCTGGATG ATAGCCGCTT TCAGATACAG CAAACCGAAA	240
	ATATCATTCG CAGCAAAACT CCCACGGGGC CGGAGCTAGA CACCAGCTAC AAAGGCTACA	300
	TGAAACTGCT GGGCGAATGC ACTAGCAGTA TAGACTCCGT GAAGAGACTG GAGCACAAC	360
20	TGAAGGAGGA AGAGGAGAGC CTTCTGGCT TTGTTAACCT GCATAGTACC GAAACCCAAA	420
	CGGCTCGTGT GATTGACCGA TGGGAGCTTC TCCAGGCCCA GGCATTGAGC AAGGAGTTGA	480
25	GGATGAAGCA GAACCTCCAG AAGTGGCAGC AGTTTAACTC AGACTTGAAC AGCATCTGGG	540
	CCTGGCTGGG GGACACGGAG GAGGAGTTGG AACAGCTCCA GCGTCTGGAA CTCAGCACTG	600
	ACATCCAGAC CATCGAGCTC CAGATCAAAA AGCTCAAGGA GCTCCAGAAA GCTGTGGACC	660
30	ACCGCAAAGC CATCATCTTC TCCATCAATC TCTGCAGCCC TGAGTTCACC CAGGCTGACA	720
	GCAAGGAGAG CCGGGACCTG CAGGATCGCT TGTSGCAGAT GAATGGGCGC TGGGACCGAG	780
35	TGTGCTCTCT GCTGGAGGAG TGGCGGGGCC TGCTGCAGGA TGCCCTGATG CAGTGCCAGG	840
	GTTTCCATGA AATGAGCCAT GCTTTCCTTC TTATGCTGGA GAACATTGAC AGAAGGAAAA	900
	ATGAAATTGT CCTATTGAT TCTAACCTTG ATGCAGAGAT ACTTCAGGAC CATCACAAC	960
40	AGCTTATGCA AATAAAGCAT GAGCTGTTGG AATCCCAACT CAGAGTAGCC TCTTTGCAAG	1020
	ACATGCTCTG CCAACTACTG GTGAATGCTG AAGGAACAGA CTGTTTAGAA GCCAAAGAAA	1080
45	AAGTCCATGT TATTGGAAAT CGGCTCAAAC TTCTCTTGAA GCAGGTCAGT COTCATATCA	1140
	AGGAACTGGA GAAGTTATTA GACGTGTCAA GTAGTCAGCA GGATTTGTCT TCCTGGTCTT	1200
	CTGCTGATGA ACTGGACACC TCAGGCTCTG TGAGTCCCAV ATCAGGAAGG AGCACCCCAA	1260
50	ACAGACAGAA AACGCCACGA GGCAAGTGTA GTCTCTCACA GCCTGGACCC TCTGTACGCA	1320
	GTCCACATAG CAGGTCCACA AAAGGTGGCT CCGATTCTTC CTTTCTGAG CCARGGCCAG	1380
55	GTCGGTCCGG CCGCGGCTTC CTGTTACAGG TCCTCCGAGC AGCTCTTCCC CTTACAGTTC	1440
	TCCTGCTCCT COTCATCGGG CTGTCCTGCC TTGTACCAAT GTCAGAGGAA GACTACAGCT	1500
	GTGCCCTCTC CAACAACCTT GCCCGGTCAT TCCACCCCAT GCTCAGATAC ACCAATGGCC	1560
60	CTCCTCCACT CTGAACCTAG CAGATGCCAT CTCAGAAGT GCTGGTAGCA TAAGGAGGAT	1620

427

CGGGTCATAA GCAATCCCAA ACTACCAACA AGAGGACCTT GATCTTGGCG AAAGCCMTCC 1680
 GTGTGGCAGC TTTAGCCTCC TCCAGATCAC ATGTGTGCAA ATTATGGCTT CAGAGGTGGA 1740
 5 AGATAAACAG TGACGGGGGA ACAAACAGAC AACAGAAGG TTTGGAAGAA ATCTGTTTG 1800
 AGACTCTGAA CCTTAGCACT AAGGAGATTG AGTAAGGACC TCCAAAGTTC CCCGGACTCA 1860
 10 TGAATTCCTG GCCCTTGGCC NATTCCTGTC ACAGCCAAGG ACTTCAGTAG ACCATCTGGG 1920
 CAGCTTTCCC ATGGTGTGTC TCCAACCATC AGATAAATGA CCTTCCCAAG CACCATGTCA 1980
 GTGTGCTACA ATCTACCAAC CAACCACTGC TGAAGAGATT TTAGAACCTT GTAACATACA 2040
 15 ATTTTAAAG GCTTATATGG CAGCTTCCTT TTTACCTTGT TTTCCTTTGG GGCATGATGT 2100
 TTTAACCTTT GCTTTAGAAG CACAAGCTGT AAATCTAAAA GGCACCTTTT TTTAGAGCTA 2160
 20 TAAAGAAAAA CTAGATGTAA TAAATAGAT CATGGAAGGC TTTATGTGAA AAAAGTTGAA 2220
 TGTATAGTA AAAAAAAG ATATTTATGT ATGTACAGTT TGCTAAAGCC AAGTTTGT 2280
 TGTATGATT TCTTTGCATT TATTATAGAT ATTATAAAAT AAAAAAAAAA AAAAAAAAC 2340
 25 TCGAGGGGGG GCGCGGTACC CAATTCGCCC TATAGTGAG 2379

30

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 1348 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

40

45

50

55

60

CCGCCTTCAC GATGCGGGCG GTCAGTGGTC CAGGTCCCTT ATTCTGCTT CTCCTCTGC 60
 TCCTGGACCC CCACAGCCCT GAGACGGGGT GTCTCTCTT ACGCAGGTTT GAGTACAAGC 120
 TCAGCTTCAA AGGCCCAAGG CTGGCATTGC CTGGGGCTGG AATACCTTC TGGAGCCATC 180
 ATGGAGGTGA GGGGCAGGGG TGGGACCGC TATGCCCAGG GTCCCTCAA GTGCTGGAG 240
 GCCTGTRACT TGGTGGGGAG TGGGTCTGTC ACAGCCATCC TCTGTCCAGG GTGGGGCAAG 300
 GCCTGGGACA GTGCCAGGCA CCCCAGGACC CTTCCAGGC TTGTCTCTG CTCCACCGCC 360
 TCAACACCCC CCACCCCTGC CCAAGTGTT TCTCTCTGC CTCTCTNNTT CCTGCCCCA 420
 GGAATTCTCT CTCTCTCTCT GCCTCTCTT GGACCCCTGC CTTCTCTTA CCTCTGACCT 480
 GTGAACACAC AGACACATGC TCACACTA AGTCCCARGC ACACMSAAAG GCAATGTGGA 540
 CCAGCACAAA CCTCCCTCT CCCGGCTCCA TCCCACGGG CTTGTGGCTG GGCATGAAAA 600

428

CTGGGGGCTA CCTGGAGGGA AGCATCTCTCA TCCCAGGTGA GTGGGCACCA GGCCTTCCCT 660
 GTATGTGTGT TGTGGGTGGA AGCAGGCATG AGAGCATCTT AGCCCATAGG TTTGTATTCA 720
 5 GGGACTTCCA AACCCAGACC TACAAAGAGT GTGTCTTCTA CCAGATCTTG TTCAAAAAG 780
 GCTTTGTGAT GATGGAACCA CACGATAGAG GGAAGTACCA AGAACATGA GGATTAGACT 840
 GGAGCGTGAA ATAGTCTAGG AGCATGGCTT CCAAAACATA TGCTGTGAGG TCTGTCCACC 900
 10 TGAGAGTTGG GCCATGGATT TAATCTGAG CCTCTTAGCA GGCAAAGCAA AGACAGAAAG 960
 CAGATCGGCT GTGGATTTCT GTCTATAAAA TGTGAGTTCT TGGCCGGGTG CGGTGGCTCA 1020
 15 CGCCTGTAAT CCGCGCGCTT TGGGAGGCCA GGGCGGATCG GTCCCGAGGT CAGGAGGTTG 1080
 GAAACCATCC TGGCCGAAT GGTGAAGCCC TGACTCTACT AGAAGTGCAA AGATTGGCTG 1140
 GGTGTGGTGG CGTCCGCTG TGGTCCAGC TTCTCGGGAG GGTGAGGCGG GAGAGTTGCT 1200
 20 TGGCCCTGGG AGGCGGAGGT TCGGTGAGC TGAGATCTG CCATTGCACT TCAGCCTGGG 1260
 CACAGAGCCA GACTCTGGCT CAAAAAATAA AAAAAAATAA ACTCGAGGGG GGCCCGTACC 1320
 25 CAATTGCGCG NATATGATCG TAAACAAT 1348

30 (2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1502 base pairs
 (B) TYPE: nucleic acid
 35 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

40 CTCAAAATAA ATAAATAAAT AAAAATTTGT ATTCCATTGA TTTGGGTAGA CACCAGGAAT 60
 GTGCATTTCT AACAAAGCTTT CCAGGCGATC CTATAGTAAG TCATCTGTGG ACTACTTTAA 120
 45 GAAACTCTTC TATAGAGAAT GGAGTTGGAT TAATAATAGG TGATTTTTTA CACTGGACTG 180
 ATTACACAAGA ACCTAAACAG TAGTCCATGA AGCTGCTCAT CTGTGGTAAC TATTTGGCCC 240
 CGTCTCACTC TGAAAGCAGC AGGAGATGTT GTTTACTTTG TTTCTATCCC CTTTGTCTGG 300
 50 AGATTAATTT TGGAAATGAA GTTTTTCTCT CTATGCCATT CCTGGTTCTT TTCCAAAGCC 360
 TCATACAAGA GGATTAGGTC ACAATGCATG CATTACCTTT TAAAAGAATG CGATATTGAT 420
 ACCGATGCTT ACTTTTTTTT TTTTNACTA CTTGTTTTAT TCCTTCCAGN AAAGTATAGC 480
 55 CCGCCTTTCT ATAGCATAGT TCTCTTAGG TGAATGATT CCTATAAGAT TTCTCATTAT 540
 TAAATCATGC ATTTTCAAG ATGGAATCAA TMTTGAATT AATCTAAGCT GATATTCTCA 600
 60 TTTGTTAGAA GAACAACCTA CATGCTAGAG AGAGAGGAGG AAATATACCC ACCACCACAC 660

	AGCCAGTTAG TATCCAGTTG GTGCTGGACT CCAGCCAGGT GTCCTGCCTC ATGCTAGTTA	720
	AATGATATAT AGAAAAGGTA AATTTTAA GAAATATTTA TTAATATATT CCTATAAAAC	780
5	ATTTTAAAGG TAACCACATA AAAATGGTTA ATTTTTCAT TCCAAAGTAA ATGCTAAGCA	840
	TGTTTATTA TGAAGCAGTA CTTCTGATTA GTATATGACA TTCTGAAGTT AATTAACTC	900
10	ATTGCACTAA ATGTGTCTTC CTTGGTATAG TGGAGGATTT GAGGATTGCA ATATAGAGTA	960
	GAGTGCTTGC TTAAGCCTGG GAGCCCATCT TTATAGCTAT TTGATGTAAG AAAAGAGACA	1020
	TGGNCCATTT CTAACTATA TAAGGTGAGT GTGTCTATTC CCAGCAGATA TAAAGGAAAA	1080
15	AGGAAACTTT TTTGATTCCT ACCTTCCAG CCTCACCTAG CCATCTTCCA GCTCAAATA	1140
	TAGAGATGTT AGTGCAAGGT CCTGGGCTCT AGGTGATCAT TTCATAAGTC CTTTACAGAT	1200
20	AAAGAAAAAG TAGTGTGTTT ATGTTTGT TTAAGTAACC CCAAAACAAA TTTATATTGT	1260
	ATTCAGCAAA ATTGGAATTC AGGTGTTTAA TTTAGAACA TGAAGTGCCT GCTGTTTTAA	1320
	GCATTGACTT GTATAAAAAG AATTGCATGT CTCCAGTAAG CTTATGGGTT TTCTCATTTT	1380
25	TAGGTATATG GCTTTTAATC ATGTAAAGTG AAACATTAGT TTTCTTGCAT TTTATTACAG	1440
	GTTCTTTGTT GCAATAAAGA TGCTGCTGAA ATTAATTGAA AAAAAAAAAA AAAAAAACTC	1500
30	GA	1502

35 (2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 1637 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

45	ATTTTCTAGC CCACAAGGAC TGAAGTTCAG ATCCAAAAGT TCACTTGCTA ATTATCTTCA	60
	CAAAAATGGA GAGACTTCTC TTAAGCCAGA AGATTTTGAT TTTACTGTAC TTTCTAAAAG	120
	GGGTATCAAG TCAAGATATA AAGACTGCAG CATGGCAGCC CTGACATCCC ATCTACAAAA	180
50	CCAAAGTAAC AATTCAAAC TGAACCTCAG GACCCGAAGC AAGTGCAAAA AGGATGTGTT	240
	TATGCCGCCA AGTAGTAGTT CAGAGTTGCA GGAGAGCAGA GGACTCTCTA ACTTTACTTC	300
55	CACTCATTTG CTTTGAAG AAGATGAGGG TGTGATGAT GTTAACCTCA GAAAGGTTAG	360
	AAAGCCCAA GGAAAGTGA CTATTTTGAA AGGAATCCCA ATTAAGAAAA CTAAAAAGG	420
60	ATGTAGGAAG AGCTGTTGAG GTTTGTTT AAGTGATACC AAAAGAGAAT CTGTGTGTAA	480

430

TAAAGCAGAT GGTGAAAGTG AACCTGTTGC ACAAAAAAGT CAGCTTGATA GAACTGTCTG 540
 CATTTCGTGAT GCTGGAGCAT GTGGTGAGAC CCTCAGTGTG ACCAGTGAAC AAAACAGCCT 600
 5 TGTAACAAAA AAAGAAAGAT CATTGAGTTC AGGATCAAAT TTTTGTCTCTG AACAAAAAAC 660
 TTCTGGCCTC ATAAACAAAT TTTGTTGAGC CAAAGACTCA GAACACAACG AGAAGTATGA 720
 10 GGTACCTTTT TTAGAATCTG AAGAAATCGG AACAAAAAGT GAAGTTGTGG AAAGGAAAGA 780
 ACATTTGTCAT ACTGACATTT TAAAACGTGG CTCTGAAATG GACAACAACCT GCTCACCAC 840
 CAGGAAAGAC TTCACTGAAG ATACCATCCC ACGGAACACA GATAGAAAGA AGGAAAACAA 900
 15 GCTGTATTTT TTCCAGCAAA TATAACAAG AAGCTCTTAG CCCCCACGA COTAAAGCCT 960
 TTAAGAAATG GACACCTCCT CGGTCACCTT TTAATCTCGT TCAAGAAACA CTTTTTCATG 1020
 ATCCATGGAA GCTTGTCTAT GCTACTATAT TTCTCAATCG GACCTCAGGC AAAATGGCAA 1080
 20 TACCTGTGCT TTGGAAGTTT CTGGAGAACT ATCCTTCAGC TGAGGTAGCA AGAAGTGCAG 1140
 ACTGGAGAGA TGTGTCAGAA CTCTTAAAC CTCTGGTCT CTACGATCTT CCGGCAAAAA 1200
 25 CCATTGTCAA GTTCTCAGAT GAATACCTGA CAAAGCAGTG GAAGTATCCA ATTGAGCTTC 1260
 ATGGGATGG TGCACCCTGA AGACCACAAA TTAAATAAAT ATCATGACTG GCTTTGGGAA 1320
 AATCATGAAA AATTAAGTCT ATCTTAACT CTGCAGCTTT CAGGCTCATC TGTATGCAT 1380
 30 AGCTTTGCAC TTCAAAAAAG CTTAATTAA TACAACCAAC CACCTTTCCA GCCATAGAGA 1440
 TTTTAATTAG CCCAACTAGA AGCCTAGTGT GTGTGCTTTC TTAATGTGTG TGCCAATGGT 1500
 35 GCATCTTTGC TACTGAATGT GTTGAACAT GTTTTGAGAT TTTTAAAAA TAAATTATTA 1560
 TTTGACAACA ATCCAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1620
 40 AAAAAAAAAA AAAAAA 1637

(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2911 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

55 GGTGGTTTTT GTTCTGCAAT AGGCGGCTTA GAGGGAGGGG CTTTTTCGCC TATACCTACT 60
 GTAGCTTCTC CACGTATGGA CCTAAAGGC TACTGCTGCT ACTACGGGGC TAGACAGTTA 120
 CTGTCTCAGC TCTAGGATGT GCGTTCTTCC ACTAGAAGCT CTCTGAGGG AGGTAATTAA 180
 60 AAAACAGTGG AATGGAAAA CAGTGTGTGA GTCATCCTGT AATATGCTCC TTGTCAACAA 240

	TGTATACATT CCTGCTAGGT GCCATATTCA TTGCTTTAAG CTCAGTCCG ATCTTACTRG	300
5	TGAAGTATTC TGCCCAATGAA GAAAACAAGT ATGATTATCT TCCAACTACT GTGAATGTGT	360
	GCTCAGAAGT GGTGAAGCTA GTTTTCTGTG TGCTTGTGTC ATTCTGTGTT ATAAAGAAAG	420
	ATCATCAAAG TAGAAATTTG AAATATGCTT CCTGGAAGGA ATTCTCTGAT TTCATGAAGT	480
10	GGTCCATTCC TGCCCTTTCTT TATTTCTCTG ATAAGTTGAT TGTCTTCTAT GTCTGTCTCT	540
	ATCTTCAACC AGCCATGGCT GTTATCTTCT CAAATTTTAG CATTATAACA ACAGCTCTTC	600
15	TATTCAGGAT AGTGCTGAAG ANGCGTCTAA ACTGGATCCA GTGGGCTTCC CTCTGACTT	660
	TATTTTGTG TATGTGGCC TTGACTGCCG GGAAGTAAAC TTTACAGCAC AACTTGGCAG	720
	GACGTGGATT TCATCAAGAT GCCTTTTTCG GCCCTTCCAA TTCCTGCTTT CTTTTCAGAA	780
20	ATGAGTGTCC CAGAAAAGAC AATTGTACAG CAAAGGAATG GACTTTTCTT GAAGCTAAAT	840
	GGAACACCAC AGCCAGAGTT TTCAGTCACA TCCGTCTTGG CATGGGCCAT GTTCTTATTA	900
25	TAGTCCAGTG TTTTATTTCT TCAATGGCTA ATATCTATAA TGAAAAGATA CTGAAGGAAG	960
	GGAACCAGCT CACTGAAGCC ATCTTCATAC AGAACAGCAA ACTCTATTTT TTTGGCATT	1020
	TGTTTAATGG GCTGACTCTG GGCCTTCAGA GGAGTAACCG TGATCAGATT AAGAACTGTG	1080
30	GATTTTTTTA TGGCCACAGT GCATTTTCAG TAGCCCTTAT TTTTGTAAGT GCATTCCAGG	1140
	GCCTTTCAGT GCCTTTCATT CTGAAGTTCC TGGATAACAT GTTCCATGTC TTGATGGCCC	1200
35	AGGTTACCAC TGTCATTATC ACAACAGTGT CTGTCTGGT CTTTGACTTC AGGCCCTCCC	1260
	TGGAATTTTT CTGGGAAGCC CCATCAGTCC TTCTCTCTAT ATTTATTTAT AATGCCAGCA	1320
	AGCCTCAAGT TCCGGAATAC GCACCTAGGC AAGAAAGGAT CCGAGATCTA AGTGCAATC	1380
40	TTTGGGAGCG TTCCAGTGGG GATGGAGAAG AACTAGAAAG ACTTACCAA CCCAAGAGTG	1440
	ATGAGTCAGA TGAAGATACT TTCTAACTGG TACCCACATA GTTTGCAGCT CTCTGAACC	1500
45	TTATTTTCAC ATTTTCAGTG TTTGTAATAT TTATCTTTTC ACTTTGATAA ACCAGAAATG	1560
	TTTCTAAATC CTAATATTCT TTGCATATAT CTAGCTACTC CCTAAATGGT TCCATCCAAG	1620
	GCTTAGAGTA CCCAAAGGCT AAGAAATTCT AAAGAACTGA TACAGGAGTA ACAATATGAA	1680
50	GAATTCATTA ATATCTCAGT ACTTGATAAA TCAGAAAGTT ATATGTGCAG ATTATTTTCC	1740
	TTGGCCTTCA AGCTTCCAAA AAAGTTGTAA TAATCATGTT AGCTATAGCT TGTATATACA	1800
55	CATAGAGATC AATTTGCCAA ATATTCACAA TCATGTAGTT CTAGTTTACA TGCCAAAGTC	1860
	TTCCCTTTTT AACATTATAA AAGCTAGGTT GTCTCTTGAA TTTTGAGGCC CTAGAGATAG	1920
	TCATTTTGCA AGTAAAGAGC AACGGGACCC TTTCTAAAAA CGTTGGTTGA AGGACCTAAA	1980
60	TACTTGCCCA TACCATAGAT TTGGGATGAT GTAGTCTGTG CTAAATATTT TGCTGAAGAA	2040

GCAGTTTCTC AGACACACACA TCTCAGAATT TTAATTTTTA GAAATTCATG GGAAATTGGA 2100
 TTTTGTAAAT AATCTTTTGA TGTPTTAAAC ATTGGTTCCC TAGTCACCAT AGTTACCACT 2160
 5 TGTATTTTAA GTCAATTTAA CAAGCCACGG TGGGGCTTTT TTCTCCTCAG TTGAGGAGA 2220
 AAAATCTTGA TGTCACTACT CCTGAATTAT TACATTTTGG AGAATAAGAG GGCATTTTAT 2280
 10 TTTATTAGTT ACTAATTCAA GCTGTGACTA TTGTATATCT TTCCAAGAGT TGAAATGCTG 2340
 GCTTCAGAAT CATACCAGAT TGTCAGTGAA GCTGATGCGT AGGAACPTTT AAAGGGATCC 2400
 TTTCAAAAGG ATCACTTAGC AAACACATGT TGACTTTTAA CTGATGTATG AATATTAATA 2460
 15 CTCTAAAAAT AGAAAGACCA GTAATATATA AGTCACTTTA CAGTGCTACT TCACACTTAA 2520
 AAGTGCATGG TATTTTTTCAT GGTATTTTGC ATGCAGCCAG TTAACCTCTG TAGATAGAGA 2580
 20 AGTCAGGTGA TAGATGATAT TAAAAATTAG CAAACAAAAG TGAATTGCTC AGGGTCATGC 2640
 AGCTGGGTGA TGATAGAAGA GTGGGCTTTA ACTGGCAGGC CTGTATGTTT ACAGACTACC 2700
 ATACTGTAAA TATGAGCTTT ATGGTGTCTT TCTCAGAAAC TTATACATTT CTGCTCTCCT 2760
 25 TTCTCCTAAG TTTCATGCAG ATGAATATAA GGTAAATATC TATTATATAA TTCAATTTGTG 2820
 ATATCCACAA TAATATGACT GGCAAGAATT GGTGGAATT TGTAAATAAA ATAATTATTA 2880
 30 AACCTAAAAA AAAAAAAAAA AAAAACTCGA G 2911

35 (2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 519 base pairs
 (B) TYPE: nucleic acid
 40 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

45 GGCAGGAGCC CCAGGCCAGC CAGGGCCAGG CCTACTTTGG CCACCCTTAA ATTAGAATGT 60
 GGGGTCAGGG GTCACAGAAA AGCCATTTCT CTGACCTAGT GTTTGGGCTC CGGGAACCTC 120
 GTGCCCAACC TTCAGACCCT GGCAGTCCTC ACTGAGGCCA TTGGCCCAAG GCCCCCCATC 180
 50 CCCCAGARACC CCGGGGAGCC GCTGTGTTGC ACGTCCACAC CTGCCACACC CTCTGCCGGG 240
 CCCCAGCCCC TCCCAACCGG GACCGTGCTG GTCCCTGGGG GTCCCTGCCCC ACCTTGCCCTT 300
 55 GGGGAGGCAT GGGCCCTCCT CTTCCACACC TGCCGGCCGT CACTCACCTC TTGCTTCTGG 360
 TCCCCAGGC CTAGCCCTTG GAAGGAGACA GGAGTCTAGG GAGGCTGAAG CCCACTCCCG 420
 60 GGGAGGCCCG TGCTCCTCCA GCGCCAGGGA CAGCAAGGAA AAGAGAAGAG AGCAGAGCAT 480

433

TTCATGGCTC TAATAAAAAA AAAAAAAAAA AAAACTCSA

519

5

(2) INFORMATION FOR SEQ ID NO: 131:

(1) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 968 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(11) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

20

TCCCTTGGG	CCCCGAAAAA	GGGGGTTGG	CCTGNCCTT	GTTTNTCCAT	GGGGCCCGCT	60
CATGCCCCAG	TACTAGCCTG	CAGTCCCAAT	GTAGCCCTC	CCTCTTCCA	GAGCCCTCM	120
AACGCCCCCG	STCANTTGTG	ATTTCAGGAG	GATTTGATGA	AGATGTTAAA	GGGAAAGTGG	180
AGAACCTTCT	CGGATTTC	AGCCTGGAA	AAACGGACCC	TGTTAGGCAA	GCACCTGCA	240
GGCCTCCCTG	TCCCTTCTT	CCCCTCCCT	TCVCCGCCC	GTGGAGACAG	CTGTTTTCAG	300
CAGGCTCTC	CGCAGGAGG	GGGCGGCTC	CTTCCCTGGC	AGCAACATCC	TTGCCCTTGT	360
CACACAAGTC	AGCCTCCATC	TGCGCAGCTC	TGTGGATGCG	CTGCTGGAGG	GCAACAGGTA	420
TGTCACCTGC	TGGTTCAGCC	CCTACCACCG	CCAGCGAAG	CTCATCCACC	CGTTCATGGT	480
TCAGCACATC	CAGCCCGCAG	CGCTCAGCCT	CCTGGCACAG	TGGAGCACCC	TCGTCCAGGA	540
GCTGGAGGCT	GGCCTGCAGC	TGGCTTTCTA	CCCCGATGCC	GTGGAGGAGT	GGCTCCAGGA	600
AAACGTGCAC	CCCAGCCTGC	AGCGGCTGCA	ARCTCTGCTG	CAGGACCTCA	GCGAGGTGTC	660
TGCCCCCCCC	CTGCCACCCA	CCAGCCCTGG	CAGGGACGTT	GCTCAGGACC	CCTGAGGGGA	720
GAGCTCATGC	CAGGGGGCTC	CTGCTGGAGG	CTGGGGGGGC	TCTGCWYTK	CWWWITGGCCT	780
GGGCAATACG	GGCCACGTGG	GGTGGTGCC	CTCTGGCCCA	GCAGTGTCTT	GGCCACACTC	840
AGTTCCCTGAG	GGCCCTGGGC	AGCCCTGGG	GGAGAGACTA	GAAAACACAG	AAGGAAGCAG	900
CACAGGGAGA	CCCGCTTTGT	GATCTGCATG	TGTGACACTG	ATTCTTTGGA	AATAAAGAGT	960
GGAAGCTC						968

50

(2) INFORMATION FOR SEQ ID NO: 132:

55

(1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1128 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

5 TGTAAAAGTT ATCAGTAATC CTAATTCTTT TCCTGGGTTT TCCTTTTGTC ACTTATTAAT 60
 CAGTTTTTTGA AAGGACGAAT GAATTTAGAG ATGTACTCTG GAGCAGTATC ATGTTAAACC 120
 AGGGGTATAT TAGAAAAATC ATCCTCATAA TCATTCTGGG AAGTTPTTCC TCCCCAAAAA 180
 AAGCCATCCT GATGGGTTTT CAAAACCAGA AAAAAGCTCT TAATGAGGAA CAGACCACTG 240
 10 GAGTACCCAT GAGCATCTCA GGAAACTGA GACCCCTCGAG AAGCCTTGAT TTCGTGCAAC 300
 CCCCCAAGGT TCAGAGCCAG CAGCCCACTG CTGTGTTGA CAGACGTGGT TTTKTGGRGA 360
 15 AAGCAGCCAG AGGCCAGGAA TTTTCAGACT CTGAGTCAC GRTVTCCAC CCAAGATTAG 420
 AGCAGAGATT AGCCATACTG ACATTTGGTA AAATCATTCT GTCTAAGCAA TGGAGGTGTG 480
 TGCAMACGTG CAGTGCCGTG TCACAGGGGA TGCAGGCAGA TCSYGGGTTT AGGATGGGGR 540
 20 AGGCCACCCG ACCCCCYTTC AYTGCTCTGC ACCTGCTCCC TCACGTGGAC ACTGTCCAGA 600
 ACTGTGGCTC TCACAGGACA GTTGCCCAAG GAGCTCATAT CTTATTGGAG ATAGGGGGTC 660
 25 GTACAGGTGA CATTGATGAG CAGTGTGAGC CGGGTGACAT GGGGGTGTC AACCAGCATC 720
 TGTCCAGGAG CTCCTCTGTC AGCGGCTCTG GCAGGTGGCC TGAGGCTCCT TTTTGAGAGA 780
 GAACTGTTTG GCCTTCCTGT CTCCTCTCCT CTGATCTGTT CTTTCTTGGA ACACCACCCA 840
 30 AGAACGTGAC CTCCTCCATC AGATTGTGAG CTCCTGGAGG GCAGGAGCTG TGTCTTTCTA 900
 TTCATCTTCC TATCCCCAGA ACCTTGACACA GATCCTGGAA TGTGTTAGGT GCTCAGTAAA 960
 35 TGTGTGTTGA ATAAATGAAT GAATGAATGA ACAAATGAAT GAATTTGCTT ACTTCAAGGC 1020
 AAAAGAACCA TGAAGCTGTA TTTGAGTTT CTATGTTATA GCAGTCAGCA AATCCTATTA 1080
 40 AATACTTTGT GTTTCCTAGC AAAAAAAAAA AAAAAAAAAA AAACCTGA 1128

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2276 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

55 CCGCGGCGTC TGACCTCATG GCGTAGAGCC TAGCAACAGC GCAGGCTCCC AGCCGAGTCC 60
 GTTATGGCCG CTGCCGTCCC GAAGAGGATG AGGGGGCCAG CACAAGCGAA ACTGCTGCCC 120
 GGGTCGGCCA TCCAAGCCCT TGTGGGGTTG GCGCGGCGCG TGGTCTTGGC GCTCCTGCTT 180
 60 GTGTCCGCCG CTCTATCCAG TTTTGTATCA CGGACTGATT CACCGAGCCC AACCGTACTT 240

	AACTCACATA TTTCTACCCC AAATGTGAAT GCTTTAACAC ATGAAAACCA AACCACACCT	300
5	TCTATTTCCC AAATCAGCAC CACCCCTCCCT CCCACGACGA GTACCAAGAA AAGTGGAGGA	360
	GCATCTGTGG TCCCTCATCC CTCGCCCTACT CCTCTGTCTC AAGAGGAAGC TGATAACCAAT	420
	GAAGATCCTA GTATAGAGGA GGAGGATCTT CTCATGCTGA ACAGTTCTCC ATCCACAGCC	480
10	AAAGACACTC TAGACAATCG CGATTATGGA GAACCAGACT ATGACTGGAC CACGGGCCCC	540
	AGGGACGACG ACGAGTCTGA TGACACCTTG GAAGAAAACA GGGGTTACAT GGAAATTGAA	600
15	CAGTCAGTGA AATCTTTTAA GATGCCATCC TCAAATATAG AAGAGGAAGA CAGCCATTTC	660
	TTTTTTCATC TTATTATTTT TGCTTTTTGC ATTGCTGTTG TTTACATTAC ATATCACAAC	720
	AAAAGGAAGA TTTTCTTCTT GGTTCAAAGC AGGAAATGGC GTGATGGCCT TTGTTCCAA	780
20	ACAGTGGAAAT ACCATCGCCT AGATCAGAAT GTTAATGAGG CAATGGCTTC TTTGAAGATT	840
	ACCAATGATT ATATTTTTTA AAGCACTGTG ATTTGAATTT GCTTATGTAA TTTTATTTC	900
25	TTGACTTTTT ATATGATATT GTGCAAATGT TTGCCATAGG CAATTGGTAC TTAAATGAGA	960
	GGTGAGTCTC TCTTTTGCCCT TGGTGCTTTG GAAATTAAAT GTCACAAACG AGTATATAAT	1020
	TTTTTAATCG TACTTTTAGA GCTGAGTTTA ATCAGGTGTC CAAATGTGA GTTAAACATT	1080
30	ACCTTATATT TACACTGTTA GTTTTTATTG TTTTAGATTT ATTATGCTTC TTCTGGAAGT	1140
	ATTAGTGATG CTACTTTTAA AAGATCCCAA ACTTGTAAT AAATTCTGAC ATATCTGTTA	1200
35	CTGCTGACTC ACATTGATTC TCCGCCATTC AAATACTATT TTTATCCAC ATTTTTTTTT	1260
	GTTCGCCAAC TGTAATGTAC AAGGATATGT GTGATAATGC TTTGGATTTC AGTAATATTT	1320
	TTTTTCTCTC CAAGAAACT GCTTTGGATA TTTTLAGATA ATTTAAACAT AATTTAGGAT	1380
40	AATGATATTG CTCATCTGA CCACAATTTT AGGTAAAACA TTAATGTGT CAGAAATCTT	1440
	GGCAACAGAG ACTCTGCAGC TTGCAGTGA CATAGATAAA ATGTTACAGA GATACTATTT	1500
45	TTTTGGTTGG AATTACTATA TTAAATTTAG AAGCAGAAAC TGGTAAATG TTAAATACAT	1560
	GTACAATTGC TTTTAGTTAG CAATTGATTG TAGCATGGGT TCCTCCAAGG TTCAAGCAA	1620
	TGGGCAGAGT TTAAATTTAT ATCAGATTCT TTTACTTCGT TTATTATTTT ACAGTAAATT	1680
50	TGAATAAATC TTAGGGGTCA TATCACTTA AATAACTCTG TACCTAGGTC TTCAAATTA	1740
	AAATTATACC TGAATGAAGT TGTTTGTATA CATAAAGCAT ATTTGTGTAC AATTACCTTT	1800
55	TTTCCCCCAG ACTTGTTTTC TTGTTTTTTC TTTTTTATGG CAACTGGAAA GTATTTACTA	1860
	TGGGATTCAT TTATGTCTGT CTTTCTATCA TAAAGAATTG ATCAATATGT AAATATGTGA	1920
	TTTGAACCAT GGTGACTTA CAAGTGTGAC TACAGCTTTT TAGAAAACAT AGCCCTAATA	1980
60	TATGTTAAGC AGGACCCGGG TGAGCCAGTG GGCTTGCCCT TTATGTAGAG CTGGAAGAG	2040

GCGGTCCATC CTGTCTCTTG GCGGGACACT GTACTTTCCT AATAGGGAAG GGAAGCACAA 2100
 TGGAAATACC CCTGAACCGT TTTATTGCAG TAATTTTTTT CATATCTGAA ACTATTATTT 2160
 AATATTTTGA ATAAGATTTT AAAAAATATA TGGCAAGAT ATAAATCTAA AAAAAAAAAA 2220
 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA 2276

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2500 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

TCCAAGCTAC GCCACTCGGG CTGGGGCGTT GGGAGCGGGA GTGCAGAGCG TGGTCGTGGC 60
 GCGCGCGGTG ACAAGAGCGA GCGKAGGAG GGGTGCCAT GCGCGGCGAG CAGTTCCAGT 120
 ACGATGACAG TGGGAACACC TTCTTCTACT TCCTCACCTC CTTCGTGGGG CTCATCGTGA 180
 TCCCGCGGAC ATACTACCTC TGGCCCCGAG ATCAGAATGC CGAGCAAAT CGATTAAAGA 240
 ATATCAGAAA AGTATATGGA AGGTGTATGT GGTACGTTTA CGGTTATTAA AACCCGAGCC 300
 AAATATTATT CCTACAGTAA AGAAAATAGT TGTGCTGCA GGATGGGCAT TGTTCCTATT 360
 CCTTGGATAT AAAGTTTCCA AACAGACCG AGAATACCA GAATACAATC CTTATGAAGT 420
 ATTAAATTG GATCCTGGAG CCAQAGTAGC AGAAATAAA AAACAATATC GTTTGCTGTC 480
 ACTTAAATAT CATCCAGATA AAGGAGTGA TGAGGTTATG TTCATGAGGA TAGCAAAGC 540
 TTATGCTGCT TTAACGGATG AAGAGTCCCG GAAAAATTG GAAGAATTTG GAAATCCAGA 600
 TGGGCCTCAA GCCACAAGCT TTGGAATTGC COTGCCAGCT TGGATAGTTG ACCAGAAAAA 660
 TTCAATTCTG GTTTTACTTG TATATGGATT GGCATTTATG GTTATCCTTC CAGTTGTGT 720
 GGGCTCTTGG TGSTATCGCT CAATACGCTA TAGTGGAGAC CAGATTCTAA TACGSACAC 780
 ACAGATTTAT ACATACTTTG TTTATAAAAC CCGAAATATG GATATGAAAC GTCTTATCAT 840
 GGTTTTGGST GGAGCTTCTG AATTIGATCC TCAGTATAAT AAAGATGCCA CAAGCAGACC 900
 AACGGATAAT ATTCTAATAC CACAGCTAAT CAGAGAAAT GGCAGCATTA ATTTAAGAA 960
 GAATGAGCCT CCACTTACCT GCCATATAG CCTGAAGGCC AGAGTTCTTT TACTGTCTCA 1020
 TCTTGCTAGA ATGAAAATTC CTGAGACCCT TGAAGAAGAT CAGCAATTCA TGCTAAAAA 1080
 GTGTCTGCC CTACTTCAAG AAATGGTTAA TGTAACTGCG CAACTAATAG TAATGGCCCG 1140

GAACCGTGAA GAAAGGGAGT TTCGTGCTCC AACTTTGGCA TCCCTAGAAA ACTGCATGAA 1200
 GCTTTCTCAG ATGGCCGTTT AGGGACTTCA GCAATTTAAG TCTGCCCTTC TGCAGCTCCC 1250
 5 TCATATGAA GAGGACAATC TTAGACGGGT TTCTAATCAT AAGAAGTATA AAATTAAAAC 1320
 TATCCAGGAT TTGGTGAGTT TAAAAGAATC AGATCGTCAC ACTCTACTGC ACTTCCTTGA 1380
 10 AGATGAAAAA TATGAAGAGG TTATGGCTGT CTTTGGGAGT TTTCCATATG TGACCATGGA 1440
 TATAAATCA CAGGTGTTAG ATGATGAAGA TAGCAACAAC ATCAGAGTAG GATCCTTAGT 1500
 TACAGTGTG GTTAAGTTGA CAAGGCAAA C AATGGCTGAA GTATTTGAAA AGGAGCAGTC 1560
 15 CATCTGTGCT GCAGAGGAAC AGCCAGCAGA AGATGGGCAG GGTGAAACTA ACAAGAACAG 1620
 GACRAAAGGA GGATGGCAAC AGAAGAGTAA AGGACCCAAG AAAACTGCTA AATCAAAAAA 1680
 AAAGAAACCT TTAACAAAAA AACCTACACC TGTGCTATTA CCACAGTCAA AGCAACAGAA 1740
 20 ACAAAGCAG GCAATGGAG TCGTTGGGAA TGAAGCTGCA GTAAAGCAAG ATGAAGAAGA 1800
 ACTTTCAGAT AAGGGCAGTG ATTCTGAAGA AGAAGAAACC AATAGAGATT CCCAAAGTGA 1860
 25 GAAAGATGAT GGTAGTGACA GAGACTCTGA TAGAGAGCAA GATGAAAAAC AAAACAAAGA 1920
 TGATGAAGCA GAGTGGCAAG AATTACAACA AAGCATACAG CGAAAAGAGA GAGCTCTATT 1980
 GGAAACCAAA TCAAAAATAA CACATCCTGT GTATAGCCTT TACTTTCTTG AGGAAAAACA 2040
 30 AGAATGTTGG TGGCTTTACA TTGCAGATAG GAAGGAGCAG ACATTAATAT CCATGCCATA 2100
 TCATGTGTGT ACCCTGAAAG ATACAGAGCA GTTAGAGCTG AAGTTTCCTG CACCAGGCAA 2160
 35 GCGTGGAAAT TATCAGTATA CTGTGTTTCT GAGATCAGAC TCCTATATGG GTTTGGATCA 2220
 GATTAACCA TTGGAAGTTK GGAAGTTTAT GAGGCTGAAG CCTGTGCCAG AAAATCACCC 2280
 ACAGTGGGAT ACAGCAATAG AGGGGGATGA AGACCAGGAG GACAGTGAGG CTTTGAAGA 2340
 40 TAGCTTTGAG GGAGGAAGAG GGAGGGAGGA AGGAAGGTGG TGGACTTAAG GCAGTTACTC 2400
 TGAATGGGA CCCACAGTGT TTTGCACCAT ATTTTGGCAA TTTTTTTTC CCGTTTITNG 2460
 45 GAAGTGTTTT CATTNAANCC CAGGAACCAT TACAGAACCG 2500

50 (2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1337 base pairs
 (B) TYPE: nucleic acid
 55 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

60 CTTCCGGTTC TCCGGGCAGC TGGCACTGCT GTAGCTTCTG CCACCTGCCA CCACCGGGCC 60

	TCTCCCTGGC GTTGGGTAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCTCTTGGG	120
5	GCCAGCGTGG CGGGCCTGGC GGCTCCCGGG TGGTGAGAGA GCCGTCCGGG AACCATGAAG	180
	GCCTCGCAGT GCTGCTGCTG TCTCAGCCAC CTCTTGGCTT' CGCTCCTCCT CCTGCTGTG	240
	CTGCCTGAAC TAAGCGGGYC COTGGMAGTC CTGCTGCAGG CAGCCGAGGC CGCGCCAGGT	300
10	CTTGGGGCTC CTGACCCTAG ACCACGGACA TTACCGCCGC TGCCACCGGG CCTACCCCT	360
	GCCCAGCAGC CGGGCCGTGG TCTGGCTGAA GCTGCGGGGC CGCGGGGCTC CGAGGGAGGC	420
15	AATGGCAGCA ACCCTGTGGC CGGGCTTGAG ACGGACGATC ACGGAGGGAA CGCCGGGGAA	480
	GGCTCGGTGG GTGGCGGCTT TGCTGTGAGC CCCAACCTG GCGACAAGCC CATGACCCAG	540
	CGGGCCCTGA CGGTGTTGAT GGTGCTGAGC CGCGCGGTGC TGGTGTACTT CGTGGTCAGG	600
20	ACGGTCAGGA TGAGAAGAAG AAACCGAAAG ACTAGGAGAT ATGGAGTTTT GGACACTAAC	660
	ATAGAAAATA TGAATTGAC ACCTTTAGAA CAGGATGATG AGGATGATGA CAACACGTTG	720
25	TTTGATGCCA ATCATCCTCG AAGATAAGAA TGTGCCTTTT GATGAAAGAA CTTTATCTTT	780
	CTACAATGAA GAGTGGAATT TCTATGTTTA AGGAATAAGA AGCCACTATA TCAATGTTGG	840
	GGGGTATTT AACTTACATA TATTTTAACA ACCTTTAATT TGCTGTTGCA ATAAATACCG	900
30	TATCCTTTTA TTATATCTTT ATATGTATAG AAGTACTCTR TTAATGGGCT CAGAGATGTT	960
	GGGGATAAAG TATACTGTAA TAATTTATCT GTTTGAAAAT TACTATPAAA CGGTGTTTTT	1020
35	TGATCGGTTT TGTTTTCTG CTTACCATAT GATTGTAAAT TGTTTTATGT ATTAATCAGT	1080
	TAATGCTAAT TATTTTGTCT GATGTCATAT GTTAAAGAGC TATAAATTCC AACACCAAC	1140
	TGGTGTGTAA AAATAATTAA AAATTTCTTT TACTGAAAGG TATTTCCCAT TTTTGTGGGG	1200
40	AAAAGAAGCC AAATTTATTA CTTGTGTGTT GGGTTTAA AATATTAAGA AATGTCTAAG	1260
	TTATGTTTTG CAAAACAATA AATATGATTT TAAATCTCT TAAAAAATA AAAAAAACC	1320
45	CCGGGGGGGG GCCCGG	1337

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

GGCAGGAGCC TGGACCGAGC AGCCACCGCC GCGTCCCTCT CTCCACGAGG CTGCGCGCTT 60

AGGAGCCCCA GCTCCGACAT GTCGCCCTCT GGTCCGCTGT GTCTTCTCAC CATCGTTGGC 120
 CTGATTCTCC CCACCAAGAG ACAGACGTTG AAAGATACCA CGTCCAGTTC TTCAGCAGAC 130
 5 TCAACTATCA TGGACATTCA GGTCCCGACA CGAGCCCCAG ATGCAGTCTA CACAGAAGTC 240
 CAGCCCCACT CTCCAACCCC AACCTGGCCT GCTGATGAAA CACCACAACC CCAGACCCAG 300
 ACCCAGCAAC TGGAAAGAAC GGATGGGCCT CTAGTGACAG ATCCAGAGAC ACACAAGAGC 360
 10 ACCAAAGCAG CTCATGCCAC TGATGACACC ACCACGCTCT CTGAGAGACC ATCCCCAAGC 420
 ACAGAGCTCC AGACAGACCC CCAGACCCCTC AAGCCATCTG GTTTTCATGA GGATGACCCC 480
 15 TTCTTCTATG ATGAACACAC CCTCCGGAAA CGGGGGCTGT TGGTCCGAGC TGTGCTGTTC 540
 ATCAGAGGCA TCATCATCCT CACCAGTGGC AAGTGCAGGC AGCTGTCCCG GTTATGCCCG 600
 AATCATTGCA GGTGAGTCCA TCAGAAACAG GAGCTGACAA CC/GCTGGGC ACCCGAAGAC 660
 20 CAAGCCCCCT GCCAGCTCAC CGTGCCGAGC CTCCTGCATC CCCTCGAAGA CCCTGSCCAG 720
 AGAGGGAAGA CACAGATGAT GAAGCTGGAG CCAGGGCTGC CGGTCCGAGT CTCCTACCTC 780
 25 CCCCCACCTT GCGCGCCCTT GAAGGCTACC TGGCGCCTTG GGGGCTGTCC CTCAGTTAT 840
 CTCCTCTGYT AAGACAAAAA GTAAAGCACT GTGCTCTTTG CAAAAAAA AAAAAAAA 900
 AAAAAAAA AAAAAAAA AAAAAAAA AAAAACTCG A 941
 30

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 654 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

GAATTCGGCA CGAGGCAGCT TGTGCTTTAA AGGAGGTGTT CAAAGCATGT CTGAGCAGAG 60
 45 ACTTTTGGGC TGTGTTTTAA TTAATACTTT AAAATAATTC ATATTTAAAA TATCATATGT 120
 TTCCATAAAG AGGAGGATGT TTAAATGCCT CCAGACTACA TTCTTTTATA TTCTTGATT 130
 50 TTACCTGGGA GTCCAAAGTT CAATTCCCAT AAAGCAAGCG TTTTATTTGT CACTTTCAAT 240
 ATACATCCGA TTGCCATGCT TAAGATGCAA TATGGGCTGC GGAAATAGGT TAACCCACAG 300
 GCTCCGAGG CCCAGTGTAG AAGGTGAGAG ATTCTGTGTA AATGATCAA ATAAAAGCAA 360
 55 GACCCTGGCC GGGTGGCGTA RCTCACCCCT GTAATCCAG CACTTTGGGA GGCCGAAGCG 420
 AGTGGATGAC GAGGTTAGGA GTTGGAGACC AGCCTGGCCA ACATCGTGAA ACCCCGTCTC 480
 60 TACTAAAAAT ACAAAAATTA GCGGGGCATG GTGGCAGGCA CCGTAATCC TAGCTAGTTG 540

GCAGGCTGAG GCAGGAGAAT CTTTGAATC TGGGACTTGG AGTTTGTGAC TGAGCTGAGA 600
 TGGGGCCACA GCACTCCAGC CTGGGTGACA GGGTGAGACT CTGTCTCAAA NAGA 654

(2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1848 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

GAAACTGGAC CGGAGAACCG GACCGAAGCG AAGCGGAACC CCGGAATGAG GCCGGACTGG 60
 AAAGCCCGAG CGGGGCCAGG CGGGCCCTCC CAAAAGCCTG CCCCTTCATC CCAGCGGAAA 120
 CCGCCGGCCC GCCCGAGCCG GCGGGCCGCT CGGATTGCAG TCGCGGGCGC GGAGGAAGAG 180
 AGACGGCTCC GGCAGCGGAA CCGCCTGAGG CTGGAGGAGG ACAAACCCGC CGTGGAGCGG 240
 TGCTTGGAGG AGCTGGTCTT CCGCGACGTC GAGAAGCAG AGGACGGGTT GCTGCGGCGT 300
 CTGCGAGGCC CGAGGGTTCA AGAACATGAA GACTCGGGTG ACTCAGAAGT GGAGAATGAA 360
 GCAAAAGGTA ATTTTCCACC TCAAAAGAAG CCACTTTGGG TGGATGAAGA AGATGAAGAT 420
 GAGGAAATGG TTGACATGAT GAACAATCGG TTTCGGAAGG ATATGATGAA AAATGCTACT 480
 GAAAGTAAAC TTTCGAAAGA CAACCTTAAA AAGAGACTTA AAGAAGAATT CCAACATGCC 540
 ATGGGAGGAG TACCTGCCTG GGCAGAGACT ACTAAGCGGA AAACATCTTC AGATGATGAA 600
 AGTGAAGAGG ATGAAGATCA TTGTGTGCAA AGGACTGGGA ATTTTCATATC CACATCAACT 660
 TCTCTTCCAA GAGCCATCTT GAACATGAAG AACTGCCAGC ATGGGAATGC TGAACGTCTT 720
 ACTGTTGCTC GGATCTCCAT CTGTGCAGTT CCATCCCGGT GCACAGATTG TGATGGTTGC 780
 TGGGATTAGA TAATGCTGTA TCACTATTTC AGGTGATGG GAAAACAAAT CCTAAAATTC 840
 AGAGCATCTA TTGGAAAAGG TTTCCAATCT TTAAGGCTTG TTTTAGTGCT AATGGCGAAG 900
 AAGTTTTAGC CACGAGTACC CACAGCAAGG TTCTTTATGT CTATGACATG CTGGCTGGAA 960
 AGTTAATTC TGTGCATCAA GTGACAGGTT TGAAAGAGAA GATAGTGAGG AGCTTTGAAG 1020
 TCTCCCCAGA TGGGTCTTTC TTGCTCATAA ATGGCATTCG TGGATATTTG CATTTGCTAG 1080
 CAATGAAGAC CAAAGAACTG ATTGGAAGCA TGAAATTTAA TGGAAAGGCTT GCAGCATCCA 1140
 CATTCCTTTC AGATAGTAAG AAAGTATACG CCTCTTCGGG GGATGGAGAA GTTTATGTTT 1200
 GGGATGTGAA CTCAAGGAAG TGCCTTAACA GATTGTGTGA TGAAGGCAGT TTATATGGAT 1260

TAAGCATTGC CACATCTAGG AATGGACAGT ATGTTGCTTG TGGTTCTAAT TGTGGAGTGG 1320
 TAAATATATA CAATCAAGAT TCTTGTCTCC AAGAAACAAA CCCAAAGCCA ATAAAAGCTA 1380
 5 TAATGAACCTT GOTTACAGGT GTTACTTCTC TGACCTTCAA TCCTACTACA GAAATCTTGG 1440
 CAATTGCTTC AGAAAAAATG AAAGAAGCAG TCAGATTGGT TCATCTTCCT TCCTGTACAG 1500
 TATTTTCAAA CTTCCTCAGTC ATTAAAAATA AGAATATTTT TCATGTTTCA ACCATGGATT 1560
 10 TTTCTCCGAG AAGTGGATAC TTTGCCTTGG GGAATGAAAA GGGCAAGGCC CTGATGTATA 1620
 GGTTCACCA TTAICTCAGAC TTCTAAAGAG ACTATTTGAA GTCCAGTTGA GTCACAAGAG 1680
 15 AAGCCTGTCT TGATATATCA TCTCAGAAAC TTTCTGAAT ATGTGATAAT ATATGGAAAA 1740
 TGATTTATAG ATCCAGCTGT GCTTAAGAGC CAGTAATGTC TTAATTAACA TGTGGCAGCT 1800
 TTTGTTTGAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTCGA 1848
 20

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1146 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

AAAAAAAAAACC CAGGGGAACN TTGGGGGCCG CTTTMMNTTC CCCCTCCAGG CCATTGGGGA 60
 35 ATTCTTCAAG TTAATCTTGC TTGCTCTTG GCCAAGAGGG CTTGTAGGGG GGAGAGACCC 120
 AGGATCATCA AGGGGTTTGA GTGCAAGCCT CACTCCAGC CTTGGCAGGC AGCCCTGTTC 180
 40 GAGAAGACCG GCTACTCTG TGGGGCGACG CTCATCGCCC CCAGATGGCT CCTGACAGCA 240
 GCCCCTGCG TCAAGCCCCG CTACATAGTT CACCTGGGGC AGCACAACCT CCAGAAGGAG 300
 GAGGGCTGTG AGCAGACCCG GACAGCCACT GAGTCCTTCC CCCACCCCGG CTTCAACAAC 360
 45 AGCCTCCCCA ACAAGACCA CCGCAATGAC ATCATGCTGG TGAAGATGGC ATCGCCAGTC 420
 TCCATCAGCT GGGCTGTGCG ACCCCTCACC CTCTCTCAC GCTGTGTAC TGCTGGCACC 480
 50 AGCTGYCTCA TTTCGGGCTG GGGCAGMACG TCCAGCCCCC AGTTAGGCTT GCTCACACC 540
 TTGSGATGCG CCAACATCAC CATCATGAG CACCAGAAGT GTGAGAAGCG CTACCCCGGC 600
 AACATCACAG ACACCATGGT GTGTCCAGC GTGCAGGAAG GGGGCAAGGA CTCCTGCCAG 660
 55 GGTGACTCCG GGGGCCCTCT GGTCTGTAAC CAGTCTCTTC AAGGCATTAT CTCCTGGGGC 720
 CAGGATCCGT GTGCGATCAC CCGAAAGCCT GGTGTCTACA CGAAAGTCTG CAAATATGTG 780
 60 GACTGGATCC ACGAGACGAT GAAGAACAAT TAGACTGGAC CCACCCACCA CAGCCCATCA 840

CCCTCCATTT CCACTTGGTG TTGTTTCTT GTTCACTCTG TTAATAAGAA ACCCTAAGCC 900
 AAGACCTCTT ACACACATTC TTGGGGCTTC CTGGACTACA GGAGATGCTG TCACTTAATA 960
 ATCAACCTGG GGTTCGAAAT CAGTGAGACC TGGATTCAAA TTCTGCCTTG AAATATTGTG 1020
 ACTCTGGGAA TGACACACCC TGGTTTGTTC TCTGTTGTAT CCCCAGCCCC AAAGACAGCT 1080
 CCTGGCCATA TATCAAGCTT TCAATAAATA TTGCTAAAT GAAAAAATA AAAAAAAAAA 1140
 ACTCGA 1146

(2) INFORMATION FOR SEQ ID NO: 190:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

ACTCCCTCAC CCAGGTCCCA GCCCTGGGAA CCACCTACCG TGAGCCCTTT TGCAGATATA 60
 GACTCATTTT ATCCTCAGAT GGTCTCTCAA GGTAGGTACT TTAGTCCCAT TTTAGAGATG 120
 AGACGATTGA GGCAGAGGG GTGNGTAAC TTGCCTGGGG GCTCAGGAGC ACAAAGGAG 180
 CCGAGGCAGG ATCTGACCCT TGTCTCTGG CCTCACTGCC CTCACCTTGC CATGACCCGA 240
 AGTTATGTCC CTACAAAGCA ATGCATGGTC CAAGGYTCTT TTTATTGTAT TTTTATTTT 300
 AAGGGTCTTG TTCAAACTG GTGTGAGCTC TGAGGAGTCC TGAACCTTGG GTGCAGCATC 360
 CTAGCATCCT GGGAGTCTTT TTCTGCCAC ACTGAGCTGG GTCCTCGAG GGGTGGGGCT 420
 GCTGTCTCTG GAAGCCTGGC AGCAGCACTG TATCGGGTTG GCTGAAGCTG ATCCCGCTGC 480
 GGTGCAGGGC TCCMGGAATC CCGTCTTGGC TGAAGGGGTT CCTGTAGCC MGGGATGTTT 540
 ATGAGGTCTC TCTGATGCCC CAGGGCAGG ACATGTGTGC GGGTGGAGAA AAGCAGGCCC 600
 TTTCACTGCC AGCTCCACTC AATTTCTATG TGGACCAAGA ACGATAAACT TAAAAATTT 660
 TTTTCTCTAA GGTATCTTCA GAATATGGTG TATTTTATG TGGAAAAGAA AAGTTATGAA 720
 GGCAGCTGTT ACTTTAAGAG AAAATTCATT AAAAGTCTC GAGGTATGAA GATGACGGCG 780
 TGCTTCTCAA TCAATTTGGC ATAAGTTGAT TGTGGCTGTA ATTTTTTTTT TTTTTTTTGT 840
 CAAGCATGTC AGACAATAAA GTCTTTGTAA AAAGRGAAAA AAAAAAAAAA AAAAAAATA 900
 ACTCGA 906

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1941 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

```

5      CTTCAGCTGA AGCCCAGGGA CGCCTTTTTC ACCCTGGGCC CCAATGCCGT CTTTCCCCCG      60
      CAGAGACTGG TCTTGGAAAC CCTCAGCAAA CTCAGCATCC AGGACAACAA TGTGGACCTG      120
15     ATTCTGGCCA CACCCCCCTT CAGCCGCCCTG GAGAAGTTGT ATAGCACTAT GGTGGCCTTC      180
      CTCAGTGACC GAAAGAAGCC GGTGTGCCCG AGATGGCTGT GGTACTGCTG GCCAACCTGG      240
20     CTCAGGGGGA CAGCCTGGCA GCTCGTGCCA TTGCAGTGEA GAAGGGCAGT ATCGGCAACC      300
      TCCTGGGCTT CCTAGAGGAC AGCCTTGGCG CCACACACTT CCAGCAGAGC CAGGCCAGCC      360
      TCCTCCACAT GCAGAAGCCA CCCTTTGAGC CAATTAGTGT GGACATGATG CGGCGGGCTG      420
25     CCGCGCGGCT GCTTGCCTTG GCCAAGGTGG ACCAGAACCA CTCAGAGTTT ACTCTGTACG      480
      AATCAGGGCT GTTGGACATC TCGGTATCAC CGTTGATGAA CTCAGTGGTT TCACAAGTCA      540
30     TTTGTGATGT ACTGTTTTTG NATTGGCCAG TCATGACAGC CGTGGGACAC CTCCCCCCCC      600
      CGTGTGTGTG TCGGTGTGTG GAGAACTTAG AAAGTACTG TTGCCCTTTA TTTATGCAAA      660
      ACCACCTCAG AATCCAGTTT ACCCTGTGCT GTCCAGCTTC TCCCTTGGGA AAAAGTCTCT      720
35     CCTGTTCTC TCTCTCCTT CCACCTCCCC TCCCTCCATC ACCTCAGGCC TTTCTGTTCC      780
      TTGTCTTCAC CTTACTCCCC TCAGGACCTT ACCCCACCTT CTTTGAAAAG ACAAAGCTCT      840
40     GCCTACATAG AAGACTTTTT TTATTTTAAC CAAAGTACT GTTGTTTACA GTGAGTTTGG      900
      GGAAAAAATA TAAATAAATA ATGGCTTTCC CAGTCTTGC ATCAACGGGA TGCCACATTT      960
      CATACACTGT TTTAATGGTA AAAAAAATA AAAAAAATAC AAAAAAAT TCTGAAGGAC      1020
45     AAAAAAGGTG ACTGCTGAAC TGTGTGTGGT TTATTGTTGT ACATTACAAA TCTTGCAGGA      1080
      GCCAAGAAGT TCGCAGTTGT GAACAGACCC TGTTCACCTG AGAGGCTGTG GCAGTAGAGT      1140
50     GTAGACCTTT TCATGTACTG TACTGTACAC CTGATACTGT AAACATACTG TAATAATAAT      1200
      GTCTCAGATG GAAACAGAAA ACGCTGGGTC AGCAGCAAGC TGTAGTTTTT AAAAAAGTTT      1260
      TTAGTTAAAC GTTGAGGAGA AAAAAAATA AGGCTTTTCC CCCAAGTAT CATGTGTGAA      1320
55     CCTACACAC COTGACCTCT TTCTCTCTC CTGATTTGTA TGAATAACCC TGAGATCACC      1380
      TCTTAGAACT GGTTTTAACC TTTAGCTGCA GCGNCTACGT CNAWCGNTGT GTATATATAT      1440
60     GACGTXGTAC ATTGCACATA CCCTTGGATC CCCACAGTTK GGTCTCTCTC CCAGCTACCC      1500
  
```

CTTTATAGTA TGACGAGTTA ACAAGTTGGT GACCTGCACA AAGCGAGACA CAGCTATTTA 1560
 ATCTCTTGCC CAGATATCGC CCTCTTTGGT GCGATGCTGT ACAGGTCTCT GTAAAAAGTC 1520
 5 CTTGCTGTCT CAGCAGCCAA TCAACTTATA GTTTATTTTT TTCTGGGTTT TTGTTTGTGT 1580
 TTGTTTTCTT TCTAATCGAG GTGTGAAAAA GTTCTAGGTT CAGTTCAAGT TCTGATGAAG 1740
 10 AAACACAATT GAGATTTTTT CAGTGATAAA ATCTGCATAT TTGTATTTCA ACAATGTAGC 1800
 TAAAACTTGA TGTAAATTCC TCCTTTTTTT CCTTTTTTGG CTTAATGAAT ATCATTTATT 1860
 CAGTATGAAA TCTTTATACT ATATGTTCCA COTGTTAAGA ATAAATGTAC ATTAATCTTT 1920
 15 GGTAAGACTT TAAAAAATAA A 1941

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2118 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

30 AAATAATAAT AANAATAAAT AAAAATWAAG TGCTTAKTGT AACTCAGCGG ACAGGGGCTCC 60
 CAGCTGCTCT GGCAGGTGGG ACACCTTCCA CCTTGCACAC AACAGGCCATG CAACAGGAC 120
 35 TGGATATGGT GGGGTAGAGT GCTTCTGGTG TGTTCACTTT AAGAAAACAT CTGCCAAGAG 180
 AGAAGAGTGC CCAGGAAAGA CCAGGAAAAT ACAAGTACAT GGCTGCTTCA TACCATATAC 240
 40 CCCAATTCTT TAAAGCAGCA AAAGGCACTT TTTTTTTCAG GCCAGAGTGA ATCTAAAACA 300
 AACCTGGCTT TGCTTACAGG GAAGCTGTCC CAGAAGGACT GAGTGATGCC TCTTGTTCCT 360
 TAAGGTCTGG AGAGTCTTTG CAAGTTTCCA ACGACATTTT CAACCAGGTG GGAGAGACCA 420
 45 GCAGTTGACG AGACAACTCA GACCCAAAAA ACGACGCCAA GGTAGTGAGT GGGTGCTTAT 480
 TTGGGAGTAG GATGATTTGA GAAAAACAGG AAGAAAAACC GGTGAGAAAG TGGCACTTTG 540
 50 GAAGTGGAAA GCTGTTTGCA AATAGCAACT CTGGCTAAAG CGAAAATGTT AATCAAGTAG 600
 AAAGTAAAAT TCAGGATCTT AGAAGCTCAT CCTTCTGATG AGAACTATTT TTTTTCCGT 660
 GAAGGAACCTA TTATTACTTT AAAAGTGAGG GTAATTTACA TATGGGGTGT ATATATTCTA 720
 55 AAAATAGTAA TAAAGTACC TTTTATAAGC AATGTTGTGT GGCTTGTAGA AGAAAGCAGG 780
 GAGGAAAAAA AGGCAGGCAA AACTAGTCTA GGTCTAGGCC CTAAAAATGA GCTTCCTTCC 840
 60 CACTTGACTG GAAACGCCCC TGTGATTTCT AGGCTGAAAA TAGGTAGGAT TTAACGAGTA 900

445

ACCTAGTTTC CTTCTGTCTC TGATTCTGA TCAGCTGATG GAGGTGCTAG TAAGAGGGGC 960
 CGATCATGCT CCCAGACGAG TCCTTTGGCC TCTTGCTCTC CATCCCAAGC CTGACTCCTT 1020
 5 CAGCAGCAGC CCCCTCCTTC TGTGTCCATC TGATGCAGGC AAGCAGGAGC AGTAAGAGGG 1080
 CATCCCATGT TCCAGTTTAC CTTCTATGGG GTGACTARGA GGTTCCTGGT AACTAGGGCA 1140
 10 GCCCARGCCC ACCAGGTTTC AAAAGCAGCT GCAAGCTTCA GAAACCCACT TCCTCCAACA 1200
 CCAGGGAGGT GGCAGAGAGC CCATCCAAAA GCCCACTGGG AGAGGCATAA GATTCTGTGC 1260
 CAGGCCCCCA GGTCCCTCTT GTGTACGTA GGCTCTGCTA CTGGCTCTG AAGTAAAGGC 1320
 15 AAANACAAAC GCGCAGGGCA GGGTGGCAGG AATAAAAAAC TCTGCACAGA AACCTTTTAA 1380
 AATAAGGAAA TTCCACCCCT CCCAATCCTT CCATGGAAGG GTGAGACCTT AATGTGATGT 1440
 AAGAGGAAGG TCTTCTCTGG CTTTCAGGGA AACAGCTGCA GCTGAAACTT AGGGCCCCAT 1500
 20 TCCAGGGCAC TTTTCACCAC AGCCAGTGCA GCGCTCCAA GTGCCACTGT CAGCCCCATC 1560
 ACTGCCAATT TCACAAAGCG GTTGGTCTTT GGCTTGGTCA GGACATCTTT TGTTCGATCT 1620
 25 TCAGGCGGCA GAAGTCCCGG AANACCGGTG CCGCAGCACC ATATCAGGCC TCTGCTGGGC 1680
 TGATGCCAGC TCAAACTCTT TGAAGTAGA GGCTCCCGTC CTCTCAGCTT GCTGTTGGGC 1740
 AGCGGCTCC CGAGCAAGTT CGGATGGGGG AAAGTGAACA AAAAGGTCTC CTSTCTGCTG 1800
 30 ATCAGTGTCT CATAGGGCAA GTCTGAGGG ATCTGGGACA ACAGGTGGTG GACCGAGGCC 1860
 ATGTACAGT CACAGTCCAG GACTTCCTGC TCGGATACA ACACAATCAC GGCTGCAAAG 1920
 35 TAAATCGGCA TCAGTGGGTG GCAGGCCAGG AAGAAGTCAT ATAACCGCAC GACGTGCCTG 1980
 AAGTCAGACA GGACATGCCC AAACCAGGTG ATGAGCCAGC TGAGGGCAA GATGCTCCCT 2040
 40 ACCTCAGCAC TCTGCATGAA GTCATGGAGC TCTGATTCA CCTGCTCAAT GATGGGCATC 2100
 AGATAGTTTA ATATATGC 2118

45

(2) INFORMATION FOR SEQ ID NO: 193:

(i). SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1538 base pairs
 50 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

55

CCGGGTTGGG CTCTGTGTCA GCAGCCGGGC GCGGCTGGG CCGGACATGG CAGCCTGTAC 60
 AGCCCGGGGG CTTGGCCGTG GGCAGCCGCT GGTGGTCCCG GTCCGTGACT GNGGCCGGT 120
 60 CGCCPAGGCC GCTCTGTGGC CGCCCGNAGC TGGAGCCTTC TCGCTAGCCT CGACCAGGAC 180

	GACCGGGAGG CACCTCTCGT CCCGAAACCG ACCAGAGGGC AAAGTGTGG ACACAGTTGG	240
5	TGTGTTTGAG GTGCCAAAC AGAATGGAAA ATATGAGACT GGGCAGCTTT TCCTTCATAG	300
	CATTTTTGGC TACCAGGTTG TCGTCCTGTT TCCCTGGCAG GGCAGACTGT RTGACCGGGA	360
	TGTGGCTTCT GCAGCTCCAG AAAAAGCAGA GAACCTTGCT GGGCATGGCT CCAAGGAGGT	420
10	GAAAGGCAAA ACTCACACTT ACTATCAGGT GCTGATTGAT GCTCGTGAAT GCCCACATAT	480
	ATTCAGAGA TCTCAGACAG AAGCTGTGAC CTTCTTGGCT AACCATGATG ACAGTCGGGC	540
15	CCTCTATGCC ATCCCAGGCT TGGACTATGT CAGCCATGAA GACATCTCC CCTACACCTC	600
	CAGTGATCAG GTTCCCATCC AACATGAACT CTTTGAAAGA TTTCTTCTGT ATGACCAGAC	660
	AAAAGCACCT CCTTTTGTGG CTGGGAGAC GCTAAGGGCC TGGCAAGAGA AGAATCACCC	720
20	CTGGCTGGAG CTCTCCGATG TTCATCGGGA AACAACAGAG AACATACGTG TCACTGTGAT	780
	CCCTTCTAC ATGGGCATGA GGAAGGCCA GAATTCACAC GTGTACTGCT GGGCCTACTG	840
25	TATCCGTTTG GAGAACCCTG ACAGTGATGT GGTACAGCTC CGGGAGCGGC ACTGGAGGAT	900
	ATTCACTCTC TCTGGCACCT TGGAGACAGT GCGAGGCCGA GGGTAGTGG GCACGGAACC	960
	AGTGTATCC AAGGAGCAGC CTGCGTTCCA GTATAGCAGC CACGTCTCGC TGCAGGCTTC	1020
30	CAGTGGGCAC ATGTGGGGCA CGTCCGCTT TGAAAGACCT GATGGCTCCC ACTTTGATGT	1080
	TCCGATTCCT CCGTTCTCCC TGGAAAGCAA TAAAGATGAG AAGACACCAC CCTCAGGCCT	1140
35	TCACTGGTAG GCCAGCTGAG GCGCCAAAGT GCCAGGCTTG GTCACCGGGA AGAACAACCTC	1200
	TCATCCACA ATTGCTGCAG AACTCTTCTC TCCCATCAT GGGCAGAGT GGGTCTCTTA	1260
	ATTTGATTGT GGGTTCTTT TTGTGGGGAG GGGTGGTATA ACTTTCTTTC AGAAGACCCA	1320
40	TGTGGGACAC CTCCAAGGT GGCCTCTCA TAAGCCCTGC CTACACCATG TTCCAGTAAA	1380
	CCTCTCCACC AAGGAACTGT GTTCAGCTGC CACAGGCTG GAGGAGTTTC CTGGCCTGTC	1440
45	ACGTGAGGTT TGATCAGTAA ACCAGTGCAS GYTTGGCCAA AAAAAAAAAA AAAAAAAAAA	1500
	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTCGA	1538

50

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:

55

(A) LENGTH: 1098 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

447

AGACCCCTGTC TCAAATAATA ATAATAATAA TAATCTTATT TTGGAGAATA AAGAGACCTS 50
 TGGATTGAG GTGCCATTG GGTAGAAAGA AAAGACGTTT ACACCGAGAA ATAGTCTGTG 120
 5 TTGCCCTGAA GGAGCAGAGG GATGCATCGC TGGAGGTGAC CTACAGTTGA AGAAGACTCA 180
 TTATGACAGA CTTTGTCCTT CTTCCTTGTG GAAAGTGTTC CCTCTGCTGC TACTGCTCAT 240
 10 GAGACTCTTC CCCCTCCCTG TCCAGGGAA CCAAAGGGCT TTNCTACCAC ACCCTTTCTT 300
 NGCCCCCCCC CTCCCATGTC TGCTGTGCCT TTGTACTCAG CAATTCTTNG TTGCTCCCA 360
 TTATCTTCCA GCGGGATACA GAGTGAATAG TTAACACAC TTAGGTCAA TAGGATCTAA 420
 15 ATTTTGTTC CTGCTCNGT GTAAAGAGGC CAGTGTGTTGT GTGTTGCAAG CAGCCTTGCA 480
 ATAGTAACTC TTCTCATTTG TTGGGATCT GGGCCMCAAG TTCCAGAATG ATACCGGAT 540
 CAGTGCAGAA GTTCATCAGG CTCTCGGACC TTAGGGCTGT TGGAGAAGGC TTCAGCAGCA 600
 20 GAACTGATGG TKAWKG/TCG TGTTCTCCAT CCTCAACTTT CTTTGCTTCG ATCATAACA 660
 AGAATACATT TGGAAAGGCA AAAAATGAAC ACTGTTGTTT ATTGCAGCCG TGTTTTGTGA 720
 25 CACAGATGCA CAGTCTGCTG TGAAGACCTT CTCTCAAGTG GSATVTGGGA GTCCATGCCA 780
 GATCATGGTG CTTCATGAGA GACTGACAGC TATCAGGGGT TGTGGCACTT AGTGAGGACT 840
 30 CTCTTCCCCC AGTGTGTGCT GATGACACAT ACACACCTGA CAATAGCTTG AGTCTTCTCT 900
 GTTCTTTTTA CTCTGTAGCC AACATACACA TGATTTAAAA CCCTTTCTAA ATATCTATCA 960
 TGGTTGATCC TTGTCCAAAT GCAGATCAG AGCTATTTGT ACTTCATTAT TATTTCCAAG 1020
 35 GCGAATAGTT GCCTTTCTTT TTGCAAAAAT AATTAAAGTT TTTGTATGTT GCAAAAAAAA 1080
 AAAAAAAA CTACGTAG 1098

40

(2) INFORMATION FOR SEQ ID NO: 195:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1001 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

55

60

GAATTCGGCA CGAGATAGCT TGCATCTCAT CCCAGTAAAA CCACTTATTT ATAACATATC 60
 AACGTATTGA CAAGGTTGAA GAGCAAGATT GTTCTGAGGT GAGATGCAA TTTCAAAGGG 120
 GTGAGCACTA ATTGTTCCAG TGATTGTTTA TTTATTTGGCT AGGACATAAT TACTCTCTTT 180
 GAGGTTACAC ATCTGCCTCC AGGTTCTGT GTGCTGTGTC CCTTGGGATC AGGCCAGGCC 240
 60 AGACTGTGAT CACTGAGATT CAAACTCCCA GARTATTCAG CAAGAGCTTT CTAGAGACCA 300

5 AGGCCAGGCC TGATCCCTGA GGGATGCATG AGAAGGCTTG GAATCTCATT CTGCTATGGT 360
 GGCTCTCTCT TGATCTTCTT GGAGTAGCAA AAACAGCAAT GTGGGCCCAA TGGTGTGGCC 420
 TAAATGATCA CAAAGGTAAA TGAGTAAAGG GCTCAGCAGA TGAGTAAGGA GCCTTGTCTT 480
 GAGAAATTAG CACTGGGCTC TGCATTCAGA AACATGTGAT AAGCATTGCC CATTGCACAT 540
 10 TGCCTTTTATT GTGTAAGGAC ATGAAATTCC AGTTTTGCAT AGCTAGTGAT GAATACCTGA 600
 AGGGAATTGC AGACATATTT TATTTTATTT TTAATTGACA GATGGAATTG TATATATTTA 660
 TCATGTACAT AATCATGCTT TAAATATGT ACATTATGGA ATGGCTAAAT CAAACTAACC 720
 15 TAGGCATTAT CTCATATAAT TGTCATTTTT GTGGCGAGAA GACTAAAAAT CTACCCCTTC 780
 AGCATTTTTA AAGAATACAA TGTGTTTTAT TAACAACAGT CACCATTGCG TACACTAGAT 840
 20 CTCTTGAAGT TCTTCTCTT ATCTAACTGA GATCTTGTA CTTTGATAA CAGCTCCCAA 900
 GCCCTTCCCC AACCCTGCT CCACCCGTGG TAACCACCAT TCTATTCTCA ACTTCTGCT 960
 25 AATCACCATT CTAGACACAG GGAAGACTCT CTACCCCTCG A 1001

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1443 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

40 ATAAACTGAA ATAGGTCATG CAAATATAAA ATATTATTTT TAAATTATTT GTCATAAGAA 60
 ACGATGGTGG CCATATTTTG CTTTAATAAT GGAAAAATG TGGTTAGCAT TCTKTGGAAG 120
 GTGGTCATCA GATAGTAGAC ATTTCTAGG ATTTATTTCT ACCTGCATAT GTGGAAATGT 180
 45 GTACTACTTT AGATTTATWT AATGGCAGCT AACTCAGAGG CATCAAAATG TGCTAATGGT 240
 GTAATATGGC CTTTGTCTTG CTGTCTGTT TTGTARGCCT TCAATCAAGC AAGGGCAGGG 300
 CCGTACAGTG AACTTGTCTT TTGSCAGAGC CCAGCGTCTG CCCCTGACCC CGTCTCCACT 360
 50 CTCTGTGTCC TGGAGGAGGA GCCCCTTGAT GCTACCCCTG ATTCACCTTC TGGGTGCTT 420
 GTACTGAACT GGAAGAGCC GTGCAATAAC GATCTGAAA TCCTTGCTTA CACCATTGAT 480
 55 CTAGGAGACA CTAGCATTAC CGTGGGCAAC ACCACCATGC ATGTTATGAA AGATCTCCTT 540
 CCAGAAACCA CCTACCGGTG AGTGCAAGGG AGTAGAAATC TGCATCAGCA CATCAGCACT 600
 60 TGGGGATCTA AGTAAACCTC TCGGGGAAAA TGACCAAGTG GATGTCTCT CCCAGCTGTT 660

449

TCTAAGAGCC CAGATGTCCA GAGTATGTCT TCACTTTGAT CCTCAGTCC AGAAGAGCTG 720
 TGA AAAAGCC AACTGGTTC AGGACTCAG TGGAGGTTT TGTGTGACT TTAAGTTGCA 730
 5 CCGTCTCTAC CCGAGAGTGG ACTCAATCC TCAATCTCT CTCTGAAGT TGTGTGCA 740
 AATTATAAAA GGGCTTTGGC AATATGTTAG CCGAAGATT TGGCTTTCTC GAAAGATTGT 900
 GCGGACNTTA ACAGTGGCTT AATGATGGT AAGATTTTA AGATTTCTTA AAGTGTGCA 960
 10 TTGGAGATAC GTTGACTTTT ATTAAAGAC CTATATTTGT TTAATGATT CTAAAAAAT 1020
 ATCTGGAGCT CAGGGGTTCA ACTGAGGAA CAGCTTTCA GATCACTGT CTATTAATA 1080
 15 AATGCCAGGT AAGCGTTGA AATTATCAA AAGCTTTCC AGTACGCA AAGTCTCA 1140
 GAGGATAGTT CTGTTATGGA GAAGATGAA TGGTTTATA GTGTAGGAC CATGCAAGG 1200
 TGAGCTTACA TTGGATAGT AAAACCTCA GACCTATT TAAAGTATT TTAATATGC 1260
 20 AGCATAAATA ATTTAATCA GTGTTAANAT GCGAGGCTA GTATATTAG CTAAATGTCA 1320
 AAGAAACTC ACATGGGAG AATGCCACT TCTCTTATA AGATAGCTT GAATATCCA 1380
 25 TTTTAGACAG ATCGAAATTG AATAGCTTTA GAAAGGCA ATGTTTCACT TTGGGAAAA 1440
 AAA 1442

30

(2) INFORMATION FOR SEQ ID NO: 197:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

GAAAAA AAAA AGTATGACCC AGTAGCTAGG CACCTGTGGC CCGGCCAGT TGACACATAA 60
 AATTAACTGT CACAGTATCA TCTTAGAAGT GAAGAGAGC CTTTATCTT GCAATGCCCC 120
 45 TCTACCACCA CCACTGACA AAGAAGATGG TGCTACTGG CATGGAGAA ATTTCTAGTT 180
 TGCTATGGCT TGTATGTGTC CCTCAAATT CAGTGTTCG CAATGTGCA GATCAAGAG 240
 50 GTGGGTCTT TAAGAGATCA CTAGGCCATG AGGATCTTC TTAGGACTGG GATGAGGCC 300
 CATAATAAAA GAGGTTTCAG GGAGCATCTT GCTATTTTC GTTCTGTATG TGAGACACA 360
 GCAAGAAAGC CCTAGTCAAC AAGTGGCAGC TCTTCACTT TAGACTTCCC ACTTCCAGA 420
 55 ACTGTGAGAA ATACATTCT GTTCCTTACA AATTAAGAG TCTCTGTAT TCTTTATAG 480
 CAGCACAAAA TGAAGATACC ATACCTGAC ACCTGACAT TCTTACAGG GTACTAAATG 540
 60 CACTGCTTTA TCTGGTCTC ACTATGCTT GCTTAATAG GAATGAAA AATTTGGATC 600

450

5 AGGGCATAGG ATGAACAACT TACTGCTAGA COTCTCAGAA TGCCACTAAT GGAAGAGATT 660
 GTATTTTCAT CATTCCTTGT CTCTTCGGAA GGTACACCGA TGGTATATA GGAATTAAT 720
 AGATGTCTAA AAGACACTTA AGTATTTGTC TAGAAATCTG GTGCATTGTC GAAAGGAAC 780
 CAAAATTGMA AATRAATTTGA AAGGGCCTAA AGCACTAATT AATCGAATT CATTAATTTT 840
 10 TAATGGTACT ACCACTCTCA AATTTAAAAT GTCATCTTAA GTTCTCTTTC CTGCGATGCG 900
 ATTTATGCTT AAAACCTGGT AAAGACTTTA ATCGATTTCG ATTCGATTAC CATGCTCTTT 960
 GTCCAGAATT ACTCCGAGAC TAATAGTCAC CTGACTTCTC GCGCTGCTTC GCGATTGCTT 1020
 15 GTCTAATTCT GGTACAAAT AAGTAACTGC CAACCTAATC TTCTTAAAA GACAGACTCA 1080
 TCTCGTCACT CTTTGTCTCA ACAATGTAAA AGCTCCGATT GTCTGCTAAA TAAAGCAGC 1140
 20 TTTCCACTGT GTATACAATA CATCCATGAT CTGTATCCAG CATCATTTTC TATTGCTCA 1200
 CTTTATACAC CACCCCCCAT GCCACATCAA ATTAAATTAT COTGATTAAT GCAATTGCAA 1260
 25 AAAAAAAAAA AAAAAAATC GA 1282

(2) INFORMATION FOR SEQ ID NO: 198:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 951 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

40 ATTTCCGAAC GAGGACTGAA GTGGGACCGG CGGCAGGTTA GAGACAGAA GGGGTATCTA 60
 TGTGGTAACT AAAGAATGTT TCTGTTTGT TAATTAATTT GTGTGTGTCG TTTTATGTT 120
 TGCTTAAGAG AATCAAAAC TGAAAAAAT GAGAATACAG GAAATGGTC TTGTTTATTT 180
 45 TTTTGTGTG TTTACAGCTT GTTAATGCTC TACTGTCTTT GTTTCAGAG AATTGTTTC 240
 ACTGCCCAGC TCGTTTGTG TCTGAGCCC TAGCCCCAGC CCACTTATA AATCTGCTT 300
 GTTTAGATGT TTGATTTTGT TCTGTTTGT ATTGTTATCT TAAGGTGTA TAACTCTGAC 360
 50 ATGCCAGACA TCAAAATTAAG CTCAAATTAA GCTCTGTTT AATGTTTAA ACACTTAAT 420
 TATATTCTAA TTGATCCCAG CCACTGATGC ATGTACTTTA GGTACTTTTG CTAAATAAGC 480
 55 ATATTAATTT TCCACATCAG GCCATCAGAT CTTGAGAACC AACGTTATC TAGAATCCG 540
 TGTCTACTAA TGTTCACCT GCATGCAGCC TTCAATTAAT TTGTATCAA ATATAAGTG 600
 ATCATTATGT AGTTTCTGGA TAAAAAAT TTGTCTGTA AGTTGTTTTG TAAATGCAAT 660

451

CTGGAATTAA TCGGACACTG TCCCTTTTGT GTTAGATGTT AGACCAAAAG AAAGGCCTTA 720
 TACTGTTAGT ATTGGAGCAC TTGAAGATA GATATTTTCA GAAAAGATGT AGGATTTAAA 780
 5 AGTTAAATTT TAAATTTTAG AAAAGATAT GATGGCAATT GGAAATAGTC ACAATGAAGT 840
 TCTTCATCCA GTAGGTGTTT AACAGTGTTA TTTTGCCACT GGTAAATGTGT AAAGTGTGAG 900
 10 TCATTTACAA TAAATGATTA TGAATTCAAA AAAAAAAAAA AAAAAACTCG A 951

(2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1740 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

TTATTATAAT AATGATGATG ATTGCAACGA AAAAACCTAC ACCGAATGTT CCATTTCTAC 60
 25 CCGGCACCCA GACACTCTCC CTAACACTGA TAACCTGAGC CCCCAGCACT GGACCGAAGA 120
 ATGCTGGCGT CTCCTGTGT ACTGGTTCAG GGTCTGCGC CCAGCCTTGT CAGGACCCCC 180
 30 TGGTGTCCAG AGCCCCCACC CTTCCCGCAA CAAGCAGCTG ATGCCCGAGT GATTCTCTAT 240
 ACATTTTCA CCTCGGCCAA TATGTCCAGG AAAACTGCTT ACTTCTCTTT TCTTGCCCTGG 300
 AGCCTTCATT GTTCACCCTT ACSTTGCAAT ATAGGAATTA ATGCTACAAA ATAAAAGTAA 360
 35 AGCTTACCTG AAAAGTGCAT AGTTTGGGGC AATGGTATCT ACATCTCCCA CTGTGGGAAA 420
 ACCAGCAAAG CATCAAACT CTCATTCTC CTGTTACCTA ATGCAGATCT GAATTATAAG 480
 40 ATGTTTATGT TTGACCATG TTTCAACAAT GGGATTTTGT TACGAATTAT CCCTTTAACT 540
 GAAACCTCA GTTTTACTGT TTACATTAT AGGAAAACAG GGATATCTTT TGAATCTAAA 600
 AATTTGATGT ACAGCATGT ATTTTGAAG TTTACATGTA AAGTCACGT ATAGGTGAAA 660
 45 TAACGTTTGT CATATTTTGA GACGTATCCT GCAGCCATGT TTTTACGTGA GTGTTTTAGT 720
 CAAAGTACAT GGTAGACAGT CTTTCACAAT AAAAGGAAAA GGATTTTTTT TCCTCCAAAT 780
 50 GTACATTTAT CAACCTAATG ATTGATTTTT TAAAAAGAG ATTTGGCCCC AGTCTGTTTT 840
 ATGAAAGTTC ATGGCCCTAA ACTGTGCTGA TTGTTTTTAA TCAAGTTATA AATTTCACAC 900
 55 CTAGATCATG TATCTACCAA CTCTCCTGCA TTTTCCAAAA GGCATTGAGC TTAATATTA 960
 GTCTTGCTTA GAGTAGGTTA TCCACTTACA TGCTGCGCTA AAGCCATGCC TTTGAAACTC 1020
 CTTGTTTTAA ACATGATATG ATTTTGTGG GCACTTTTCA AAAAGAAAC AAACAACAA 1080
 60 AAATCGACCC TTAATTATT ACTTGCAACT CAACAGATCT CCCTGCCGTA CTGCCTTTTC 1140

452

CAGGAACCTTT ACTTCAGGGC TGTCCAGATT GCAGTTGTGC CCCGTGTATG TCGATCTACT 1200
 TCACAGAGTC TTTGGAAGCC AGCAGTCGTG CCGTCCGTAT ACTGTCCACT CATTTTATGT 1260
 5 AGATTTCGTA TCCTCAGCAG CCAGTGTAA CACCACTGTC ACGTAGTTAN CAGATTGATC 1320
 TTTTATGTAT TTAAGTAAT CCATACTATG ATTTGGTTTT TCCCTGCACC ATTAATTCTG 1380
 10 GCATCAGATC AGTTTTTGTG TTGTGAAGTT CTACTGTGGT TTGACCCAAG ACCACAACCA 1440
 TGAGACCCCTG AAGTAAAGAT AAGGTACACA TACATTATTT GAGTAACTGT TTCCTTGGGG 1500
 GCCAATCTGT GTATGCTTTT AGAAGTTTAC AGAATGCTTT TATTTTGTG TATAACAAAC 1560
 15 AGTCTGTCTAT TATTTCTGT TGATAACCA TTGGACAGA GTGAGGACGT TTGCCCTGTT 1620
 ATCTCCTAGT GCTAACATA CACTCCAGTC ATGAGCCGGG CTTTACAAAT AAAGCACTTT 1680
 20 TGATGACTCA MAAAAAAAAA AAAAAAAMC YCGGGGGGGG GCGCGTAACC CATTNNCCC 1740

25 (2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1707 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

35 GCTTATAGAA GGGAGAGGAG CGAACATGCC AGCGCGTTGG CCGTTTTGCT GTGTCTCTGT 60
 GACCATGGTG GTGGCGCTGC TCATCGTTTG CGACGTTCCC TCAGCCTCTG CCCAAAGAAA 120
 GAAGGAGATG GTGTTATCTG AAAAGGTTAG TCAGCTGATG GAATGGACTA ACAAAGACC 180
 40 TGTAATAAGA ATGAATGGAG ACAAGTTCCG TCGCCTTGTG AAAGCCCCAC CGAGAAATTA 240
 CTCGGTTATC GTCATGTTCA CTGCTCTCCA ACTGCATAGA CAGTGTGTGG TTTGCAAGCA 300
 45 AGCTGATGAA GAATTCAGA TCCTGGCAAA CTCCTGGCGA TACTCCAGTG CATTACCAA 360
 CAGGATATTT TTTGCCATGG TGGATTTTGA TGAAGGCTCT GATGTATTC AGATGCTAAA 420
 CATGAATTCA GCTCCAACTT TCATCAACTT TCCTGCAAAA GGGAAACCCA ACCGGGTGA 480
 50 TACATATGAG TTACAGGTGC GGGGTTTTTC AGCTGAGCAG ATTGCCCGGT GGATCGCGA 540
 CAGAACTGAT GTCAATATTA GAGTGATTAG ACCCCCAAAT TATGCTGGTC CCCTTATGTT 600
 55 GGGATTGCTT TTGGCTGTTA TTGGTGGACT TGTGTATCTT CGAAGAGTAA TATGGAATTT 660
 CTCTTTAATA AAAGTGGATG GGCTTTTGCA GCTTTGTGTT TTGTGCTTGC TATGACATCT 720
 60 GGTCAAATGT CGAACCATAT AAGAGGACCA CCATATGCCC ATAGAAATCC CCACACGGCA 780

453

CATGTGAATT ATATCCATGG AAGCACTCAA GCCCAGTTTG TAGCTGAAAC ACACATTGTT 840
CTTCTGTTTA ATGGTGGAGT TACCTTAGCA ATGGTGCTTT TATGTGAAGC TGCTACCTCT 900
5 GACATGGATA TTGGAAGCG AAAGATAATG TGTGTGGCTG GTATTGGACT TGTGTATTA 960
TTCTTCAGTT GGATGCTCTC TATTTTAGA TCTAAATATC ATGGCTACCC ATACAGCTTT 1020
10 CTGATGAGTT AAAAAGCTCC CAGAGATATA TAGACACTGG AGTACTGGAA ATTGAAAAAC 1080
GAAAATCGTG TGTGTTTGA AAGAAGAATG CAACTTGAT ATTTTGATTT ACCTCTTTT 1140
TTCAAGTGAT TTAATAGTT AATCATTTAA CCAAGAAGA TGTGTAGTGC CTTAACAAGC 1200
15 AATCTCTGT CAAAATCTGA GGTATTTGAA AATAATTATC CTCTTAACCT TCTCTTCCCA 1260
GTGAACTTTA TGAACATTT AATTTAGTAC AATTAAGTAT ATTATAAAAA TTGTAAACT 1320
20 ACTACTTTGT TTAGTTAGA ACAAAGCTCA AACTACTTT AGTTAACTTG GTCATCTGAT 1380
TTTATATGCC CTTATCCAAA GATGGGGAAA GTAAGTCTG ACCAGGTGTT CCCACATATG 1440
CCTGTTACAG ATAACATACAT TAGGAATTCA TTCTTAGCTT CTTGATCTTT GTGTGGATGT 1500
25 GTATACTTTA CCGATCTTTC CTTTIGAGTA GAGAAATTAT GTGTGTCATG TGGTCTTCTG 1560
AAAATGGAAC ACCATTCTTC AGAGCACAGG TCTAGCCCTC AGCAAGACAG TTGTTTCTCC 1620
30 TCCTCCTTGC ATATTTCTTA CTGAAATACA GTCTGTCTA TGATTGTTTT TGTMTTGTG 1680
TTTTTTGAG ATCACGATAC TGGGCTC 1707

35

(2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 779 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

50

55

60

CTGTCCCAG TGTTCAGG TAATGACTTG GCACTCCAGA GAAAGTTTCA TRCTGTTGCG 60
TGTGGTGGCT CCAAGCCAAG CACCTGGCAT GCAGGTCAGC CCTTCCCAGC GGGCGTGGCG 120
TGTCTCTCTT CACAGATGCC ACGTTGCAGC CCCAAGGCCT CACCATTTTG CTTTTTTAG 180
AAACCCATT TCTTGGTCAAT TTATAAGCT GCTTTATAGA TATCTTTGAT CCTGGCATGC 240
CTTGGTTTCC TCTCCCTTCC CTCTTCCAA TCCTGGTTTC CTAACCTCCT CTGTAGTAA 300
TTCTCAACTC AACTCAAAGT CCCAAGAATT TGGAAATGTA GGATGCTGTG CGGGAGCTC 360
GAGGCTGAGG CATAATCACT GCTTCGGTTC TGCTCATCAG GGGACAGCT CCCTTACTCA 420
TGGCAGCCAT GTTGTATTGT CACAGAGCCC CCCGAATACT CTGTCTATAG TGACACACTG 480

454

TAGGTGTCAT AAATTTTAAAG AACCTGCTT TTAAGTACTA TTTATAGGTT TTTCTGTTAT 540
 ACTTGCAACC TAGTTTTTAA ATACATGAGG ATTTTATGAA AGCTTTTATAC AGACATTTAT 600
 5 AGGAAACTCA TTCTTTGATT TTAGGTGCCA TTAAATTGA TAACACTTAC TTTATAAAAA 660
 GATGCTTTTT GTCTGGATAG ACCCTTATAG TTTAAATAT CTTCATATAT TGCCATTTGA 720
 10 TCAAATAAAT TTCTTACTTA GAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAACTCGA 779

15 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1617 base pairs
 (B) TYPE: nucleic acid
 20 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

25 GGCACAGCTT TCTGTCTCTT CTTGGCTCCC TCTCTTCTC TGCTCCCTCT GCCTTCCCAG 60
 TGCATAAAGT CTCTGTGGCT CCCGGAAGTT GTTGGCAATG CCTATTTTTT GCCTTTCCCC 120
 CGCGTTCTCT AACTAACTA TTAAAGGTC TGCGGTGCGA AATGGTTTGA CTAAACGTAG 180
 30 GATGGGACTT AAGTTGAACG GCAGATATAT TTCACTGATC CTGCGGTGTC AAATAGCGTA 240
 TCTGGTGCAG GCCGTGAGAG CAGCGGGCAA GTGCGATCGG GTCTTCAAGG GCTTTTCGGA 300
 35 CTGTTTGCTC AAGCTGGGCG ACACATGGCC AACTACCGCG AGCCTGGGAC GACAAGACGA 360
 ACATCAAGAC CGTGTGCACA TACTGGGAGG ATTTCCACAG CTGCACGGTC ACAGCCCTTA 420
 CGGATTGCCA GGAAGGGGCG AAAGATATGT GGGATAAAGT GAGAAAAGAA TCCAAAAACC 480
 40 TCAACATCCA AGGCAGCTTA TTGGAAGTCT GCGGCAGCGG CAACGGGGCG GCGGGGTCCC 540
 TGCTCCCGGC GTTCCCGGTG CTCCTGGTGT CTCTCTGGC AGCTTTAGCG ACCTGGCTTT 600
 45 CCTTCTGAGC GTGGGGCCAG CTCCTCCCGC GCGCCACCC AACTCACTC CATGCTCCCG 660
 GAAATCGAGA GGAAGATCCA TTAGTTCTTT GGGGACGTTG TGATTCTCTG TGATGCTGAA 720
 AACACTCATA TAGGATTGTG GGAAATCCTG ATTCTCTTTT TTATTTCGTT TGATTTCCTG 780
 50 TGTTTTATTT GCCAAATGTT ACCAATCAGT GAGCAAGCAA GCACAGCCAA AATCGGACCT 840
 CAGCTTTAGT CCGTCTTCAC ACACAAATAA GAAAACGGCA AACCCACCCC ATTTTTTAAAT 900
 55 TTTATTATTA TTAATTTTTT TTGTTGGCAA AAGAATCTCA GGAACGGCCC TGGGCACCTA 960
 CTATATTAAT CATGCTAGTA ACATGAAAAA TGATGGGCTC CTCCTAATAG GAAGCGGAGG 1020
 AGAGGAGAAG GCCACGGGAA TGAATTCAAG AGAGATGTCC ACCGACGAAA CATACGGTGA 1080
 60

455

ATAATTCACG CTCACGTGGT TCTTCCACAG TATCTTGTMT TGATCATTTT CACTGCCACAT 1140
 TTCTCCTCRA GAAAAGCGAA AGGACAGACT GTTGGCTTTG TGTTTGGAGG ATAGGAGGGA 1200
 5 GAGAGGGAAG GGGCTGAGGA AATCTCTGGG GTAAGAGTAA AGGCTTCCAG AAGACATGCT 1260
 GCTATGGTCA CTGAGGGGTT AGCTTTATCT GCTGTGTGTT ATGCATCCGT CCAAGTTTAC 1320
 TGCCTTTATT TTCCCTCCTC CTTCTTGTMT TAGCTGTAC ACACACAGTA ATACCTGAAT 1380
 10 ATCCAACCGT ATAGATCACA AGGGGGGGAT GTTAAATGTT AATCTAAAAT ATAGCTAAAA 1440
 AAAGATTTTG ACATRAAAGA GCCTTGATTT TAAAAAAGAG AGAGAGAGAG ATGTAATTTA 1500
 15 AAAAGTTTAT TATAAATTAA ATTCAGCAAA AAAAGATTTG CTACAAAGTA TAGAGAAGTA 1560
 TAAATAAAA GTTATTGTTT GAAAAAAGAA AAAAAAAGAAW CTCGACCGCA AGGGAAT 1617

20

(2) INFORMATION FOR SEQ ID NO: 203:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1974 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

GAATTCGGCA CGAGGCTGAG GGAGCTGCAG CGCAGCAGAG TATCTGACGG CGCCAGGTTG 60
 CGTAGGTGGG GCACGAGGAG TTTTCCCGGC AGCGAGGAGG TCCTGAGCAG CATGCCCGG 120
 35 AGGAGCGGCT TCCCTGCTGC CGGCTCTGG CTCTGGAGCA TCCTCCTGTG CTTGCTGGCA 180
 CTGCGGGGGG AGGCGGGGCC GCGCAGGAG GAGAGCCTGT ACCTATGGAT CGATGCTCAC 240
 40 CAGCCAAGAG TACTCATAGG AATTGAAGAA GATATCCTGA TTGTTTCAGA GCGGAAATG 300
 GCACCTTTTA CACATGATTT CAGAAAAGCG CAACAGAGAA TGCCAGCTAT TCCTGTCAAT 360
 ATCCATTCCA TGAATTTTAC CTGGCAAGCT GCAGGGCAGG CAGAATACTT CTATGAATTC 420
 45 CTGTCTTTCG GCTCCCTGGA TAAAGGCTC ATGGCAGATC CAACCGTCAA TGTCCTCTG 480
 CTGGGAACAG TGCTCACA GGCATCAGTT GTTCAAGTTG GTTCCCATG TCTTGGAAAA 540
 50 CAGGATGGGG TGGCAGCATT TGAAGTGGAT GTGATTGTTA TGAATCTGA AGGCAACACC 600
 ATTCTCCAAA CACCTCAAAA TGCTATCTTC TTAAAAACAT GTCAACAAGC TGAGTGGCCA 660
 GGCGGGTGCC GAAATGGAGG CTTTGTGAAT GAAAGACGCA TCTGGGAGTG TCCTGATGGG 720
 55 TTCCACGGAC CTCACTGTGA GAAAGCCCTT TGTACCCAC GATGTATGAA TGCTGGACTT 780
 TGTGTGACTC CTGTTTCTG CATCTGCCCA CTTGGATTCT ATGGAGTGAA CTGTGACAAA 840
 60 GCAAACTGCT CAACCACTG CTTAATGGA GGGACCTGTT TCTACCCCTG AAAATGTATT 900

TSCCCTCCAG GACTAGAGGG AGAGCAGTGT GAAATCAGCA AATGCCCCACA ACCCTGTGGA 960
 AATGGAGGTA AATGCATTGG TAAAGCCAA TGTAACTXTT CCAAAGGTTA CCAGGGAGAC 1020
 5 CTCTGTTCAA AGCCTGTCTG CGAGCCTGGC TGTGGTGCAC ATGGAACCTG CCATGAACCC 1080
 AACAAATGCC AATGTCRAGA AGGTTGGCAT GGAAGACACT GCAATAAAG GTACGAAGCC 1140
 10 AGCCTCATAC ATGCCCTGAG GCCAGCAGGC GCCAGCTCA GGCAGCACAC GCCTTCACTT 1200
 AAAAAGGCCG AGGAGCGCGG GGATCCACCT GAATCCAATT ACATCTGGTG AACTCCGACA 1260
 TCTGAAACGT TTAAAGTTAC ACCAAGTCA TAGCCTTTGT TAACCTTTCA TGTGTTGAAT 1320
 15 GTTCAAATAA TGTTCATTAC ACTTAAGAAT ACTGGCCTGA ATTTTATTAG CTTCAATTATA 1380
 AATCACTGAG CTGATATTTA CTCTTCCTTT TAAGTTTTCT AAGTACGTCT GTAGCATGAT 1440
 20 GGTATAGATT TTCTTGTTTC AGTGCTTTGG GACAGATTTT ATATTATGTC AATTGATCAG 1500
 GTTAAATTT TCAGTGTGTA GTTGGCAGAT ATTTTCAAAA TTACAATGCA TTTATGGTGT 1560
 CTGGGGCCAG GGGAACATCA GAAAGGTTAA ATTGGGCCAA AATGCGTAAG TCACAAGAAT 1620
 25 TTGGATGGTG CAGTTAATGT TGAAGTTACG GCATTTTACA TTTTATTGTC AGATATTTAG 1680
 ATGTTTGTTA CATTTTAAA AATTGCTCTT AATTTTAAA CTCTCAATAC AATATATTTT 1740
 30 GACCTTACCA TTATTCCAGA GATTCAGTAT TAAAAAATA AAAATTACAC TGTGGTAGTG 1800
 GCATTTAAAC AATATAATAT ATTCTAAACA CAATGAAATA GGAATATAA TGTATGAACT 1860
 TTTTGCAATG CCTTGAAGCA ATATAATATA TTGTAACAA AACACAGCTC TTACCTAATA 1920
 35 AACATTTTAT ACTGTTTGTA TGTATAAAAT AAAGGTGCTG CTTTAGTTTT CTGA 1974

40

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1057 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

50

CGGCCTTCCG GGGCAACCGT TCGTCCCAAC NCGGGAAGG GTCCTGGAGN CGGGAAGTAG 60
 GAGCCTCCGA AGTCCAAGGG CGGAGCGCCC TTGCTAATA AGCCATCAG AACGTGAGAC 120
 55 GCTCCGGTGG GNCGGTGCCG TCGAGCGCGG GGTGGACTCT GGTGACTTG GCTGGCGGGA 180
 TCAAGTGCAG CTGCTTCAGG CTGAGGTGGC AGATAGTGAG CGCTGGTGGC GGAGTTAAAG 240
 60 TYAAGCAGG AGACTAATWA TGAATAGCCG AGCGGGATTC TCACACCTAG ACCGTCCGGA 300

60

457

GCGGGTTCTC AAGTTAGGGG AGAGTTTCGA GAAGCAGCCG CGCTCCGCTT CCACACTGTG 360
 CGCTATGACT TCAAACCTGC TTCTATTGAC ACTTCTTCTG AAGGATACCT TGAGATTGGC 420
 5 GAAGKTGAAC AGKTGACCAT WACTCTGCCM AATATAGAAA GTTGAAGGAA GCAGTAAAT 480
 TCAGTATCGT AAAGAACAC ACACAACA ATGTGGAATT CAGCCAGGAC TCCCAATCTT 540
 GTAAAACATT CTCATCTGA ACATAAGATG TCCCCAGCAT CTCCAATAGA TGATATCGAA 600
 10 AGAGAACTGA ACCGAGAACC TACTCTAATG GACCAGATGA GTACTTGTGA TAGTTCATCA 660
 GATTCCAAAA GTTCATCATC TTCAAGTAGT GAGGATAGTT CTAGTGACTC AGAAGATGAA 720
 15 GATTGCAAAAT CCTCTACTTC TGATACAGGG NAATTGTGTC TCAGGACATC CTACCATGAC 780
 ACAGTACAGG ATTCTTGATA TAGATGCCAG TCATAATAGA TTTCCAGACA ACAGTGGCCT 840
 TCTGATGAAT ACTTTAGAA ATGATTTGCA GCTGAGTGAA TCAGGAAGTG ACAGTGATGA 900
 20 CTGAAGAAAT ATTTAGCTAT AAATAAAAT TTATACAGCA TGTATAATTT ATTTTGTATT 960
 AACAAATAAA ATTCTTAACA CTGAGGGAAA TATGTCTTAA CTTTGTATGA TAAAAGAAAT 1020
 25 TAAATTTGAT TCAGAAAAAA AAAAAAAAAA AACTCGA 1057

30 (2) INFORMATION FOR SEQ ID NO: 205:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 721 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

40 GAATTCGGCA CGAGTCATCC CTCCTCCTCT TTCACTCCCT TACTCTTACT CTGTTTTTTG 60
 TGCTCCAGAC AGACAGACCC TACCTCTTTT GCTTCTTTTT TGTGTGTTG TTTTGAGATG 120
 GAGTGTGGCT CTTGTTGCCC AGGCTGGAGT GCAGTGGCGC AATCTGGGT CACCACAACC 180
 45 TCTGCCTCCC GGGTCAAGC AATTCTCCTG CCTCAGCCTC CCGAGAAGCT GGGGATTACA 240
 GGCATGCGCC ACCACACCCA GCTNAATTTT ATATTTTTAG TAGAGATGGT GTTCTCTCAT 300
 50 GTTGGTCAGG CTGGCCTCAA ACTCCCAACC TCAGGTGATN CCGCCTGCTT TGGCCTCCCC 360
 AAAGTGCTGG GATTACAGGC GTGAGCCACT GCGCCAGCC TCTTTTGCTC CTTTATACTC 420
 ATTAACCTAC GCCTGTAATC CTTGTTTTG GAGGCCAAG TGAGAAGGT GTTTGAGGCC 480
 55 AAGAGTTTGA GACTAGCCTG GGCAACACAG CAAGATGCCA TCTTTATAAT AAAAATAAAA 540
 ATAAAAATCA ATTAGCTGG CATGGTGAA CGCACCTGTA GTCCAGCCA ATTGAGAGGC 600
 60 TGAAGTGGGA GGATCATTTA GCCCAGGAGT TGAGGTTGCA GTGAGCCATG ATCATGTAC 660

TACACTCAGC CTGGGCAATA GAGGGACATG TTGTCTCTAA AAAAAAAAAA AAAAAACTCG 720

A 721

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2465 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

CCACCATTTA TCCAAGTAA GAGGAGTTAC AGGCAGTTCA GAAAATTGTT TCTATTACTG 60

AACGTGCTTT AAAACTCGTT TCAGACAGTT TGTCTGAACA TGAGAAGAAC AAGAACAAG 120

AGGAGATGA TAAGAAAGAG GGAGGTAAAG ACAGAGCTTT GAAAGGAGTT TTGCGAGTGG 180

GAGTATTGGC AAAAGGATTA CTTCTCCGAG GAGATAGAAA TGTCAACCTT GTTTTGCTGT 240

GCTCAGAGAA ACCTTCAAAG ACATTATTAA GCGGTATTGC AGAAACCTTA CCCAAACAGC 300

TTGCTGTTAT AAGCCCTGAG AAGTATGACA TAAATGTGC TGTATCTGAA GCGGCAATAA 360

TTTGAATTC ATGTGTGGAA CCCAAATGC AAGTCACTAT CACACTGACA TCTCCAATTA 420

TTCCAGAGA GAACATGAGG GAAGGAGATG TAACCTCGGG TATGGTGAAA GACCCACCGG 480

ACGTCTTGA CAGGCAAAAA TGCCTGACG CTCTGGCTGC TCTACGCCAC GCTAAGTGGT 540

TCCAGGCTAG AGCTAATGGT CTGCAGTCTT GTGTGATTAT CATACGCATT CTTCCAGACC 600

TCTGTACGG AGTTCCAAGT TGGTCTGATT TTCCAAGCTG GGCTATGGAG TTAGTAGTAG 660

AGAAAGCAAT CAGCAGTCTT TCTAGCCCTC AGAGCCCTGG GGATGCACTG AGAAGAGTTT 720

TTGAATGCAT TTCTTCAGGG ATTATTCTTA AAGGTAGTCC TGGACTTCTG GATCCTTGTG 780

AAAAGGATCC CTTTGATACC TTGGCAACAA TGACTGACCA GCAGCGTGAA GACATCACAT 840

CCAGTGACA GTTTGCATG AGACTCCTTG CATTCGCCCA GATACACAAA GTTCTAGGCA 900

TGGATCCATT ACCGCAATG AGCCAACGTT TTAACATCCA CAACAACAGG AAACGAAGAA 960

GAGATAGTGA TGGAGTTGAT GGATTTGAAG CTGAGGGGAA AAAAGACAAA AAAGATTATG 1020

ATAACTTTTA AAAAGTGTCT GTAATCTTC AGTGTAAAAA AAACAGATGC CCATTTGTTG 1080

GCTGTTTTTC ATTGATAATA ATGTCTACAT TGAAAAATTT ATCAAGAATT TAAAGGATTT 1140

CATGGAAGAA CCAAGTTTTT CTATGATATT AAAAAATGTA CAGTGTTAGG TATTATTTGA 1200

ATGGAAGAC ACCCAAAAA AAAAATGTGC TCCGACTAGG GGGAAACAG TAGTTCCGAT 1260

459

TTTTTCGGCAT TATTTTATT TTATTTTCTG GTTGGGCTAG CTCCCCCCC TATTTTGTG 1320
 TCTTTTATTA ACTAGTGCAT TGTCTTATTA AATCTTCACT GTATTTAATG CAGGATGTGT 1380
 5 CCTTCAGTTG CTCTGTGTAT TTTGATATTT TAATTTAGAG GTTTTGTGTTG CTTTTTGACA 1440
 CTAGTTGTAA GTTACTTTGT TATAGATGGT ATCCTTTACC CCTTCTTAAT ATTTTACAGC 1500
 10 AGTACGTTTT TTTGTAACCT GAGACTGCAG AGTTTGTGTTT TCTATATCTG AAGGATTACA 1560
 ACACAAAAAG TTATCCTGCC ATTCGAGTGC TCAGAACTGA ATGTTTCTGC AGATCTGTG 1620
 GCATTTGTCT CTAGTGTGAT ATATAAGCT GTAATTAAGA CAGAGTTCTG TTAATCTAAT 1680
 15 CAAGTTTGCT GTTAGTTGTG CATTAGCAGT AATAAAGCTA ATATATACTA TATGGTCTTG 1740
 CAACAGTTTT AAACCTCTG CATAATTGAT AATAAAAAATG CATGACATTC TGTGTTTAA 1800
 TAGACTTTTA AAATCATAAT TTTAGGTTTA ACACGTAGAT CTTGTACAG TGTACTTTTT 1860
 20 GACATAGCAA GGCCAAAAAT AACTTTCTGA ATATTTTTTTT CTTGTGTATA AGTGGAAAGG 1920
 GCATTTTCA CATATAAGTG GGCTAACCA TATTTTCRAA AGAACTTCAT CATTGTACAA 1980
 25 CTAACACAG TAACTAGCCC TTAATTATGG TGACAGTCC TTATTGGTGT GTGTGAGATT 2040
 ACTCTAGCAA CTATTACAGT ATAACACAGA TGATCTTCTC CACACACCCC ATCACCACAGA 2100
 TAATTTACAG TTCTGTTAAC AGTGAGGTTG ATAAAGTATT ACTGATAAAA AATTATCTAA 2160
 30 GGAAAAAAC AGAAAATTAT TTGGTGTGGC CATCTTACCT GCTTATGTCT CCTACACAAA 2220
 GCTAAATATT CTAGCAGTGA TGTAAATGAAA AATTACATCT TACTGTTGAT ATATGTATGC 2280
 35 TCTGGTACAC AGATGTCAAT TTGTTGTCAC AGCACTACAG TGAAATACAC AAAAAATGAA 2340
 ATTCATATAA TGACTTAAAT GTATTATATG TTAGAATTGA CAACATAAAC TACTTTTGCT 2400
 40 TTGAAATGAT GTATGCTTCA GTAAATCAT ATTCAAATTT AAAAAAAAAA AAAAAAAAAA 2460
 CTCGA 2465

45

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1480 base pairs
 50 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

55 GAATTCGGCA CGAGCTCAAG CTGGCAGGTG GTCGGGGGAG CGGCGGAGA GGAGCTGCCG 60
 GGAGTTCGTG CCCTGCAGGA CATGACACCA GTGGCATATC ACGGCCATGG GGTCTCAGCA 120
 60 TTCCGCTGCT GCTCGCCCTT CTTCTGCGAG GCGAAAGCAA GAAGATGACA GGGACGGTTT 180

460

5 GCTGGCTGAA CGAGAGCAGG AAGAAGCCAT TGCTCAGTTC CCATATGTGG AATTCACCGG 240
GAGAGATAGC ATCACCCTGTG TCACGTGCCA GGGGACAGGC TACATTCCAA CAGAGCAAGT 300
AAATCAGTTG GTGGCTTTGA TCCCACACAG TGATCAGAGA TTGGCCCTTC AGCGAACTAA 360
GCAATATGTC CTCTGTCCA TCCTGCTTTG TCTCCTGGCA TCTGTTTGG TGGTTTCTT 420
10 CCTGTTCCG CATTCACTCC TTGTGGATGA TGACGGCATC AAAGTGGTGA AAGTCACATT 480
TAATAAGCAA GACTCCCTTG TAATTCTCAC CATCATGGCC ACCCTGAAPA TCAGGAAGTC 540
CAACTTCTAC ACSGTGGCAG TGACCAGCCT GTCCAGCCAG ATTCACTACA TGAACACACT 600
15 GGTGAATTTT ACCGGGAAGG CCGAGATGGG ACGACCGTTT TCCTATGTGT ACTTCTCTG 660
CACGGTACCT GAGATCCTGG TGCACAACAT AGTGATCTTC ATGCCAACTT CAGTGAAGAT 720
20 TTCATACATT GGCCTCATGA CCCAGAGCTC CTTGGAGACA CATCACTATG TGGATTGTGG 780
AGGAAATTCC ACAGCTATTT AACAACTGCT ATTGGTTCTT CCACACAGCG CCTGTAGAAG 840
AGAGCACAGC ATATGTTCCC AAGGCCTGAG TTCTGGACCT ACCCCCACGT GGTGTAAGCA 900
25 GAGGAGGAAT TGGTTCACCT AACTCCCGC AAACATCCTC CTGCCACTTA GGAGGAAACA 960
CCTCCCTATG GTACCAATTA TGTTCCTCAG AACCCAGCAGA ATCAGTGCCT AGCCTGTGCC 1020
30 CAGCAAATAG TTGGCACTCA ATAAAGATTT GCAGAATTTA ATACAGATCT TTTCAGCTGT 1080
TCTTAGGGCA TTATAAATGG AAATCATAAC GTGGTTCTAG GTTATCAAAC CATGGAGTGA 1140
TGTGGAGCTA GGATTGTGAG TGACCTGCAG GCCATTATCA GTGCCTCATC TGTGCAGAG 1200
35 TCGCAGCAGA GAGGGACCAT CCAAATACCT AAGAGAAAAC AGACCTAGTC AGGATATGAA 1260
TTTGTTCAG CTGTTCCCAA AGGCCTGGGA GCTTTTGTAA AAGAAAGAAA AAAGTGTGTT 1320
40 GCCTTTTTTT TTTTTAGAA AGTTAGAATT GTTTTTACCA AGAGTCTATG TGGCGCTTGA 1380
TTCACCCCTC ATCCATTGGC TGGAAACATGG ATTGGGGATT TGATAGAAAA ATAAACCCTG 1440
45 CTTTGTATTC AAAAAAAAAA AAAAAAWAAA AAAAAGTCTGA 1480

(2) INFORMATION FOR SEQ ID NO: 208:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 872 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

CAGTATTTCC CTCAGTACTG TAAGCAAAG TGGTATGTTT TTCTTTCTTT ATGTCTACTC

60

461

TGTCTCTGT GGCCTTCTGG TGTACCCCTC TCTCCCTAGC CATTCACTGT CTCTAGTCAC 120
 CTCCTAGTA GCTAGTGCTC TCTAAGTTTT TATTTAATTA GAACCACTCC ATTTCATTT 180
 5 CAAGGTAGGT CAATGGGGGG AAAAGCCTCA TGATTTAAAC TGAAGTTAAC AACACAGCTT 240
 TTAAATGAA AACTCATACT CCAACTTCTA AAGTATATTT GAGCTGATTT GTTTCCAAAA 300
 CAAAGATATG CTGTACCTAA AACTGCTAAA ACAAAAATAT AAAGACAAGG ACTAGGTGAT 360
 10 TAAGGGGAGA GAAAAATCAT VTCTTTTCCA GGAAACCTTT GCTAAAATAA GCRAAACTTG 420
 ANTCTATGCT TCATGGAAAC TGACACAAAG AAAAGAACT GATGGATTGC ACAGGCCTTG 480
 15 TTATAGAAAT AGATCTATAA AAAGATCTGT CCACAGGAAA TATACACCTT CTCTGGTTC 540
 TGAACITCAA TGGGGATTTG TCACCTAGGT CTCCATCTAT AGGAATACCT TCACATACCT 600
 ATCTATTGAT GCACATATTC TGAAACAGG TACATACAAA ATTACAACAA AGGAAAAAAA 660
 20 TTCTATTGAA CACTTAAAAA TAGAAACAGG CCAGGCACGG TGGCTCATGC TGTAAATCCCA 720
 ACAATTTGGG AGGCTGAGGC TGGTGGATCA CCGAGGTCA GGAGTGTGAG ACCAGCTTGG 780
 25 CCAACATGGT GAAACCCCGT CACTACTTAA AATACAAAAA AATTAGCCT GTGTGGTGGC 840
 ACACTCNTAC AATCCNGGCT GACTCGGGAA AN 872

30

(2) INFORMATION FOR SEQ ID NO: 209:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1779 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

AATTGCCAAG ACTGCACAAA ATTACAGTGC TAATGTATAT GGTTCAGTT CACATAAAGA 60
 CAAAAGCATC TGTTATGAAA TGAGTAGTAA TATTGGGTGG TTGATTTGTT CTTAGCAGAC 120
 45 TTGGCTTCAT WTGGTCTTG AGATAAAATG GCCAGCATAA ATGCTGTTTA TATTCACGTT 180
 TTCCTAGGTG TGTGTGTGCA GGCACAGCA CCATGCCCTT GGTGTAGTCA GTGCCGAAS 240
 50 GGGTCTGTTT CTTCTTGAGC CTGCCGTCAG GGATGGTCTC CTTTTAAAGC AGGTGTGTG 300
 CAGCATTCAG TACACTGAAG GTAAGCTAAA CCATCAACAT CTCTGGTGT TTAAGATGTT 360
 ATTTTATTGG AACAACTGAC AAATGAGGGA TGTTAGCTTT GTGGCAGAAT TCCCTGCATG 420
 55 TGTGATAACT GATCTTGTTT TATTTTITGG CATTGCAACT GTGGCATAGT TACAATTTCT 480
 GTTTGKTCAT CACATTTAAA ATTGGRAGAG AACGCCCTTG AKGGATAGAG CGCCTTCAGK 540
 60 GTACTGTTTC TTATTAACCT TACTTTTTTT AAATCAACTT CCTATAGACT TTATATACAT 600

TTTGTTAAAT ATAGTTCCTA GTGACATAGA AACGATGCCGT AGTTTTCAAT TACTAATTAC 660
AAATGTTGAG GCCTAATTCT GAAAGTCCTC ATATTTAAAG GCTAGACAAC GTAATGAAAT 720
5 TTTTAACTAT TTGTATGTCA TTTTGAAAGT GTACTGCTTT ATGGTAAAG TGTTTTTCAT 780
TTGTTTCATTG TTTTCATTAT TTGTGATCAT GTTGTCTTC AATACAGGCA TAAACCTTCC 840
10 ACTCTTGAAC AAAGCAGCTG CTTTTTAAAA GCGGTAATG CTTCTTTACC TTTTATTTCT 900
TTTGTAATG AAGCTTTTCT TTAAGAATGT GACTTTAAAG TGTGTCTAT TGCATAAAAC 960
AGTTGACACT CACTTATTGT AAAGTGAAGA TTGTTCTACT GCATGTGAAG TGGACCATGC 1020
15 AGATTTCTGT ATGTTCTCAG TATGCATCAC TAGATAATAA AGTCTTTTGT GAACAAGGCA 1080
TTGTAGCCA TTTTAAAG TTTTGTCTT CAGTGCTGGT AAGTCAGGTA AACCATAAAT 1140
20 AGTTAAAGC AACCTTTTGT TTTTTCCTG AAAGTTTTTA ATTGAAAGTA TTATTAGTTA 1200
AAGATGTAAA CCTAGCCAAA ATTACCAGTT TATTAATAAT TAGGATCCTA ATTATTTCAA 1260
AAAATCCTAC AAATATTGTC AGCTTTCAGT GTAGTGACAT TATTCCTGTA GGTATATGGG 1320
25 TATAATTGAG GATTAACTA ATGTTCTGC TATTTTCTCA CTTTCTCTT TGATGGTGCG 1380
GAAAGAGAAA AAGGAAAACG GGGCACAGGC CATTCGACGC CTTCTCCAAG GGTCTGATT 1440
30 TGCTGAGACA CCAGCTTCAC CTTCTTAACA AGGCACCTAA TTACAACAAG CATGCACATT 1500
TTGGTGCAAT CAAGAATGGA AAATCAGAAT AGCAGCATTG ATTCTTCTGG TGCAGCTCAG 1560
TGGAAGATGA TGACAACCAG AAGACATGAG CTAAGGTAA GGGACTGTTT TGAAGAACCT 1620
35 TTCCATTAG TGATCAAGAT ATGGAAGCTG ATTTCTGAAA ATGCTCAGTG TGTACTCTAA 1680
TTATTTATGG TACCATTGTA ATTGTAACCT GCATTTTAGC AGTGCATGTT TCTAATTGAC 1740
40 TTACTGGGAA ACTGAATAAA ATATGCCTCT TATTATCAA 1779

45 (2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2110 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

55 GCGGCGGCTG CAGCCCGGAG CTGAGCTAGC COTCCGACCC GAGCCGTCCG AGCCGGGGAA 60
GCGGCGGCGT GCTGCCGCTC GTGGCGGCCA GAGGAGAGGA GAGGCAGCAG CATGGCGAGT 120
60 GTCTGTCCC GACGCCTTGG AAGCGGTCC CTCCTGGGAG CCGGGGTGTT GGGACCCAGT 180

	GCCTCGGAGG GGCCTCGGCT GCGCCACGCT CGGAGCCACT GCTAGAAGGG GCGGCTCCCC	240
	AGCCTTTGAC CACCTCTGAT GACACCGGCT GCGAGGAGCA GCGCAAGGAA GTCCCTAAGG	300
5	CTCCGAGCAC CTGGGGCCTT CAGCAGGTGG CCTTTMAGCC TGGGCAGAAG GTTTATGTGT	360
	GGTACGGGGG TCAGAGTGC ACAGGACTGG TGGWGCAGCA CAGCTGGATG GAGGGTCAGG	420
10	TGACCGTCTG GCTGCTGGAG CAGAAGCTGC AGGTCTGCTG CAGGGTCGAG GAGGTGTGGC	480
	TGCGAGAGCT GCAGGGGCCC TGTCCCGAGG CACCACCGCT GGAGCCCGGA GCGCAGGCCC	540
	TGCGCTACAG GCGGCTCTCC AGGAACATCG ATGTCCCAAA GAGGAAGTCG GACGCATGGA	600
15	AATGGATGAG ATGATGGCGG CCATGGTGCT GACGTCCCTG TCCTGCAGCC CTGTTGTACA	660
	GAGTCCCTCC GGGACCGAGG CCAACTCTC TCGTTCCCGT GCGGCTTGG ACCCATGGAA	720
20	GGAGAGTGGT GACATCTGGG ACAGCGGCAN CAGCACTACC AGCGGTCACT GGAGTGGGAG	780
	CAGTGGTGTG TCCACCGGCT CGCCCCCGCA CCCCCAGGCC AGCCCCAAGT ATTTGGGGGA	840
	TGCTTTTGGT TCTCCCAAA CTGATCATGG CTMTGAGACC GATCCTGACC CTMTCTGCT	900
25	GGACGAACCA GCTCCACGAA AAAGAAAGAA CTCTGTGAAG GTGATGTACA AGTGCCTGTG	960
	GCCAAACTGT GGCAAAGTTC TGCGCTCCAT TGTGGGCATC AAACGACACG TCAAAGCGCT	1020
30	CCATCTGGGG GACACAGTGG ACTCTGATCA GTTCAAGCGG GAGGAGGATT TCTACTACAC	1080
	AGAGGTGCAG CTGAAGGAGG AATCTGCTGC TGCTGCTGCT GCTGCTGCCG CAGACCCCCA	1140
	GTCCCTGGGA CTCCACCTC CGAGCCAGCT CCCACCCCCA GCATGACTGG CCTGCCCTCTG	1200
35	TCTGCTCTTC CACCACCTCT GCACAAAGCC CAGTCTCCG GCGCAGAACA TCCTGGCCCCG	1260
	GAGTCCCTCC TGCCCTCAGG GGCTCTCAGC AAGTCAGCTC CTGGGTCCCT CTGGCACATT	1320
40	CAGGCAGATC ATGCATACCA GGCTCTGCCA TCCTTCCAGA TCCAGTCTC ACCACACATC	1380
	TACACCACTG TCAGCTGGGC TGCTGCCCCC TCGCGCGGCT GCTCTCTMTG TCGGCTCCGG	1440
	AGCGGTCGCG TAAGCTTCAG CGAAGCCCCA GCAGCCAGCA CCTGCGATGA AATCTCATCT	1500
45	GATCGTCACT TCTCCACCCC GGGCCGAGG TGCTGCCAGG AAAGCCCGAG GCGAGGCTAA	1560
	GAAGTGCCGC AAGTGTATGG CATCGAGCAC CGGGACCACT GGTGCACGGC CTGCCGCTGG	1620
50	AAGAAGGCTT GCCAGCGCTT TCTGGACTGA GCTGTGCTGC AGGTTCTACT CTGTTCTGG	1680
	CCCTGCCGGC AGCCACTGAC AAGAGGCCAG TGTGTACCA GCGCTCAGCA GAAACCGAAA	1740
	GAGAAAGAAC GGAAACAGG AGTTTGGGCT CTGTTGGCTA AGGTGTAACT CTAAAGCAA	1800
55	TTTCTCCCA TTGTGCGAAC ATTTTATTTT TTAACAAAAA GAAACAAAAA TATTTTCC	1860
	CCTAAATAG GAGAGAGCCA AAAGTACCA AGGCTATTC GAGTGAACC ACTGACCAA	1920
60	GAATTAATTA CCTCCGTTT CCCACATCCC CACTCTCTAG GGGATTAGCT TGTGCGTGTG	1980

464

AAAAGAAGGA ACAGCTCGTT CTGCTTCCTG CTGAGTCGGT GAATTCCTTG CTTTCTAAAC 2040

TCTTCCAGAA AGGACTGTGA GCAAGATGAA TTTACTTTTC TTAACAAAAA AAAAAAAA 2100

5 AAAAACTCGA 2110

10 (2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 938 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

20 GGCACAGGAA AAAAAAGAAA AAAGAAAAAA GAAAAAGTT TTTGTACCCA CAGATTAGCA 60

TTTTCTGAT GTTTGAAAAA AGTTTAAGCT ATGTCCTAAT TTAACAAATGA GCACAAACTA 120

25 CTTAACAGAT GTCTGTTCCTC TCTTCTCTTA CTTAAATTAT CTTTATTTTC ACCATCACCT 180

CCCAGTGCCG AACACCTGAN CTCTGTGTTT TGTGGTTGGA TCCTGGGTTC CCAAGTTCTT 240

ATTTGGTCAG TCCCTGGCCT GTGGGGGGGT CTCAGGAAGT GGCATGCTCT TCAMGRAGGA 300

30 TCGTTTCATYT CCAGTATAAC CAWTTTGTTA ATAATAGTTG ATAATTCCTA GCTTTTACCA 360

GATGATTTT GACTTATTTT TCCTCCTTTC ACCTGTTCAA AGCTAACATA TCTCGGTCAG 420

35 TTGGGAGAGG GTGGGGGATT TGAGAATGTG AGGAGGAGTG GGGTTAGANT GGGTTTGCTT 480

ATCTGGGCAA GGAAAGAGTT CCTAGTCGAT TGGGCACAAT GACAAATGA TTCCATGGAT 540

AGAATCGTCC CATGTTGCTG GAACACCTCA CGTGTGTGTA ACGCCTTAAA TTCTGCCAT 600

40 CCCTTCTCTG ATTCCCCACC TCCTGTAGT TTCCACAGGA TTTATCTCTC TGTACCCCCG 660

TCCTCCAAC TACTCTGTC AGCCTCTCCT CCATCCCTTA CTTCCCTTCT AAATTCAGG 720

45 AGATGACCTC ACTTTGCAAA GCAAAATTGA GCCACCAAT TGTAGCTCTC CTCGGTGGAA 780

ACTGCATCTG TGCTCATCCC TGCACCTTCT TGCAGAAAGC CGCCCCCTCA GGCCAAGATG 840

AGTGCCTGGC CCCCATGGGA GACTCAGACA CTTTGACCCC TTGTGACTTC AGCATCTCCC 900

50 TCTTTAAAGA TTCTCTCCA ACATTCAGTC GTGCTCGA 938

55 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1351 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

5	AGGCTGGACT AACCATAGAG AACCAAGGAGA GAAAGAAAGA TTAAAGAGAC TGAGTAAATAT	60
	TTTTTGACAG ATCATTTAAG AAACAGAGTA ATTTTTTTTT TCTCCAAAAG GGCATGGGTT	120
10	TTTTTTTTGT TTGTTTTT CTCTATTGG CACTTTCTAG GGATTGGTCT ATAAATTTTT	180
	TGAAAGATCA TAGGATAAAT TTCTTTGTAG CAACCTCCTA TTTTAGTGTT TATGTTAGGG	240
	GATCCCCARG TGTCCCTGCT GATACGCCAT TAGGGCCACT TCTCAGCCTC TGGCTACATC	300
15	ATAATGCTTT TTTTCTATC TTGCCAAGT TTCCAGAAAA TTKAKGTTTT CTAATTTTAA	360
	AAAAATGGT TGTGGAGATG GGATGGGACC TCTTTATAAG CCTGAAAAT AAGTGATTN	420
20	TTTAAAGTGC TATCTCTGTA TAAACCTGAT TCTCACTTTT TTCTGTAGAC AACAGTTTTT	480
	TATAATATAT CTATTTTGTG TGGACATTAT TTCCTTTTAA CCAATACTGA AATTCCATAG	540
	TGTAWACTTT CTCCACATTT TCTTTGATTA ATACTTCTT AAAATAGACA CTTGGATTGG	600
25	CACCAGCTGT CACCAATAAA GCTGCCCTGA ACATTGTCAA TCAATCCTGT TAACCAATTT	660
	GAGAATTTTT CTGGAATGCT TAGTTAGGGA TGAAATGCT GGGTTATAGG TATGAGTATG	720
30	CTGATATAC TTTCTCCAG AATGTCTACA CCTGTGTGTA CACCACATCT CCAGAGATAG	780
	GGGAATCTTA TGTCCCTGCT AACTGCTCTC GTTATTTAAT TTTGTGACAT TTGCGCCCGC	840
	CGCGGCCCCC TGCCCCAAC ACACACATGG TATAAAGTGG TAGTTTCTTG TTTAAATTG	900
35	AACCTTTGAA TGATTTGAAT TTGGGCATTT CTTTGTATCC TGAGTTATTT TGCTTTCCCG	960
	TTATGTGAAT ATCCTTTTCC TATGCTTTAA CTACTTTTCT AATTGTGCCC TTTTTTNGGT	1020
40	TATCAAATTC CAGGCCATG TCTATTCCAT CGTCACTTTT GGGTATTGGA AACATCTTTC	1080
	CATTCTGTAG CTTGTCTGTT GAACATAAAT CTTGATTTTT ATGTAATCAG ATTTTTCTCC	1140
	TTACGGTTAT GTTCTTGGAA TTTTATTTAA GAAATCTTTT TCTATCCTGA GACCACAAAA	1200
45	ATGTCCCCAC CATTTCTTC TGTTTCATAG TTTTGCCTTG TATGTTTAAAT CCTTTAAGGC	1260
	ATGTGTAGTT CATTTTATAT GGTGTGAAAT AGTTCTTATT CATTTATTCA ACACATATTG	1320
50	GTGGAGTGCC TGCTGATGGT AGTACTCTTC AGAGTACTTT GTATATATTT GTGAACACAT	1380
	ATTCTTGCCC TGGAAGCTTA TGTTGTCTTT CAAGGTAGAT CONTACTCGG TTTCCACCTG	1440
	TTTTCTTCAG CCTCAGGAT GAATTCACA ATTTTACACA TAGCACCAGT TAAGGAATAG	1500
55	GCTTTATTGG AGAAAAGGAA GCCTTATTAG ACCAGCATCA GCAAAAAAAA A	1561

60 (2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 997 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

10 AGAGAGTCCT CAACAGAACC TAATCATGCT GGCACCTAA TTTCTACTT CTAGCCTCCA 60
 GAAGTGAGAG AACATAAACT CCAGTTGTTT AAGCTACCCA GTTCTAGGTA TTGTTTATTA 120
 TAGCCCAAGC TAAGTCAGGT GGAAAGGCGG AATATTTTG AAGAGATCA TTTCTACAAA 180
 15 AACAGAGTTG TTCTAAATGA AATGGCCAGA TATTTCTCT TTTCTACTT AGTATTTATG 240
 AAAGTTTCAT TAAACACCAC TTGGCCAGCA CCCAGGCTG CTTCTTTGAG AACGGCAAC 300
 20 AAAAGCAAAT GATTTGAGGA ACAAAGAGT GGACACAGAG CTTCTGAGA GATGGCTCA 360
 TCTTCTGAGA TGATCTTCTG AGATCATCAA TTTTCTGCAG CTTCTCTCT ACTCCAAATG 420
 TAGTAGATAA GAGCAAAGAC ACTTCCTGAT CTTGTGGAAA AAGCTGGAGC TCTGCTGATG 480
 25 GAGAGGCTGA CACTGGGACC AACAGAAGGC CGGACATTA TTGTTGGAG CCTTCTGGA 540
 CCTGGGCCCT CTTGAGGCTT TGTACCTTGC ACTCCGCTG CTTCTCTAGC AACTGGTAA 600
 30 CTGAAGTTAG GTATTTGAAG AGATAATTTG CCCCCACCA AATTTACTT AAAGAAAAA 660
 GGAAACCACT AAATCCACT TGACAAACCA GTTTGTTGAG TTTGACTTT TCGAAATTTG 720
 AAACTTCTC TTGGCACCA TATGATCTG TTACATTAG GTTCTCAT TCTAAGATAC 780
 35 ACAGCTAGGT CTACCAGCTG CCACTGGTCA AGAATGAAG AACCTCTGAG AGAGGATCA 840
 GTTTCTAATA ACCTAACAGT TTTCTTTGGS TATTACAAA AAAAAAAA TTAGATAAA 900
 40 ATGTCAGTG CATGCAGGA AGTACAGATA TGAATGAGA AATTTGCTT ACAAATGCA 960
 GATTTGTTG TTAATAAAT TGATTGGGAT CACTCGA 997

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1496 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

GAATTCGGCA CGAGTGACCA CAGATATCTT TGGCTTTGAG CTTCTAGCA ATGCTGTCCA 60
 CTATGTTTTT TTAATCCAT TGACATCTCA TGAATCCCA AATTTAGGCG CTTTTCATC 120

TTTTCATCT TTGTCATAGC TTCATCAGCG ACGATGGAGG TCACTTCAGC ACTATCCGGA 130
 GCGGCCTCAC GGACAGATCR GTGAATTTC TTTTCCTTTT TCTTGATGTA CCGGATGTC 240
 5 GACTCGTTAA CATTGAGCTC ATGGCCAACA GCACTGTAAC TCATGCCCTGA TTGGAGCTTA 300
 TCCAACACGC GGAMTTTCTC CGTAAGGSAM ATCAMGGTCT TCTTTCGCTT AGGAACACTG 360
 GGCARARCTT AARCACTACG CTTGGGGGCC ATTTTAGAAA GCAAAACCCAC CCACAAAAAG 420
 10 CAGAAAAAAA ACTGTCAGTA AACAGACTGN NGANAGGACT CTTTGTTTAC AGCAGAGGAG 480
 CTGGGACTAG AAGGCGGGCC TTCTCCCCAG TTCAAACCTC AGCTGGGAAC CTTACCTCCG 540
 15 CCAACTCCAA ATTTTCACCC TCTGCGCATG CCGGGGAAAS AAACCCCCAG AACAGTACCG 600
 TGATGATTGA TTTTACGGTT ACAAATACAT TTTAGCAAGT AAGTGAATTT GGCATTACGA 660
 ATTAATGATT AATGAAGGTC ACCTGTATTT CCATAGATAT GTAATTTTAT TTAAGCAGGT 720
 20 TTATTATATT AAGGCGGSEA GGCAGCGCCG AAGACTACAA GTTCCAGCAT GCACCGCGTC 780
 CCGGCGGGTT CGGGCTCCCA GCGAGGCTT CAGGGACGCC AGCCCGGAGG CATGGGCGG 840
 25 AAGTGTCTGA GGGCAACCAC GTAGTACTCT CTGCGCATGT GCAAAGCGCT GTCGGGGGCC 900
 GCCCTAGCTG CCGTCGCGGC CGCCGGGGCT CTATGGTCTC TCCCTAGAGC TTTGCCGTTG 960
 GAGGCGGCTG CTGCGGTCTT GTGAGTTTGA CCAGCGTCCA GCGGCAGCAA CATGGAGGAA 1020
 30 TTGACTCCG AAGACTTCTC TACGTCGGAG GAGGACGAGG ACTACGTGCC GTCGGGTGAG 1080
 CGATTCCGCC TGAGGCGAGA AGCGAATGCG CCCCCCGCAC GCCTCACGTG AGGCGCGCTC 1140
 35 TGCCCCCGCG GCGTCTGCC CTGTGGCCCA GGTGGTCCAG GGGGGCTCCT GTTCTCGAGC 1200
 GTCCGCTCCC TCAGGCCCCC CATCTCTGGC CGCTCCGGCC CGAGGCGTGT GCGCGTGGCG 1260
 GTTCTGTGCT CCCCTCCCGT TGGGCAGCTC CGCCCGCCGC CCCCTCTTGC AGCGCGGGAA 1320
 40 CCGCACATGG ACACGGCCCC TTGTGGCTAG GGACGCTCGT CCGTCAGCCC CGAACGACAA 1380
 CGCTGCTTCA GAAGTCGGGG CGGCAATTCG AGCCTTGGAA GTTTTCTTCA GCGCTGGCCC 1440
 45 GAGAGAGCTG CTGGCCAACA ACCCGTCCAA GATAGAGCTG TCCGNTCTCC GNCCTG 1496

50 (2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1308 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

60 TTGGCANCNG GGAGAGGGAA AGAGGAGGAA ATGGGGTTTG AGGACCATGG CTTACCTTTC 30

	CTGCCCTTTGA CCCATCAGAC CCCATTTCCT CCTCTTTCCC TCTCCCCGCT GCCAAAAAAA	120
5	AAAAAAAAGG AACGTTTTAT CATGAATCAA CAGGCTTTCA GTCCTTATCA AAGAGAGATG	180
	TGGAAGAGAC TAAAGAAACC ACCCTTTGTT CCCAACTCCA CTTTACCCAT ATTTTATGCA	240
	ACACAAACAC TGTCTTTTGG GGTCCCTTTC TTACAGATGG ACCTCTTGAG AAGAATTATC	300
10	GTATTCACAG TTTTtagccc TCAGGTtacc AAGATAAATA TATGTATATA TAACCTTTAT	360
	TATTGCTATA TCTTTGTGGA TAATACATTC AGGTGGTGCT GGGTGATTTA TTATAATCTG	420
15	AACCTAGGTA TATCCTTTGG TCTCCACAG TCATGTTGAG GTGGGCTCCC TGGTATGGTA	480
	AAAAGCCAGG TATAATGTAA CTTCAACCCA GCCTTTGTAC TAAGCTCTTG ATAGTGGATA	540
	TACTCTTTTA AGTTTAGCCC CAATATAGGG TAATGGAAAT TTCCTGCCCT CTGGGTTCCC	600
20	CATTTTtact ATTAAGAAGA CCAGTGATAA TTTAATAATG CCACCAACTC TGGCTTAGTT	660
	AAGTGAGAGT GTGAAGTGTG TGGCAAGAGA GCTTCACACC TCACTAGGTG CAGAGAGCCC	720
25	AGGCCTTATG TTAAATCAT GCACTTGAAA AGCAAACCTT AATCTGCAAA GACAGCAGCA	780
	AGCATATAC GGTCACTCTG AATGATCCCT TTGAAATTTT TTTTGTGTTT GTTGTTTTAA	840
	ATCAAGCCTG AGGCTGGTGA ACAGTAGCTA CACACCCATA TTGTGTGTTT TGTGAATGCT	900
30	AGCTCTCTTG AATTGGATA TTGGTTATTT TTTATAGAGT GTAAACCAAG TTTTATATTC	960
	TGCAATGCGA ACAGGTACCT ATCTGTTTCT AAATAAACT GTTTACATTC ATTATGGGGT	1020
35	ATGTATGACC TTCATTTTCC AAGAAATAGA ACTCTAGCTT AGAATTATGG ATGCTCTAAA	1080
	ATGTCAGAAT GGGAACTCTC CTCGAAGTTC TCCCAAACCTC AGAGACAGCA CTGCCTTCTC	1140
	CTAAATGATT ATTCTTTTCT CCTGTTTTTC TGGTATTTTC TAGGCATCCT TCTCACCACA	1200
40	GCCATAACCC TTTTtactt CCATTAGGCC GTATAACTGG NCGGACNGCT GGTCCGTATA	1260
	TAATACTGGT WCCAACAMAG GGGTTCTGGA TGTACACMAG GTTATCTT	1308

45

(2) INFORMATION FOR SEQ ID NO: 216:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1705 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

	TGCCCATGGA AGCGCTAGAA GGTTTAGATT TTGAAACAGC AAAGAAGGAT TTCCTTGGAT	60
60	CTGGAGACCC CAAAGAAACA AAGATGCTAA TCACCAAACA GGCTGACTGG GCCAGAAATA	120

	TCACAGGAGCC CAAAGCCGCC GTGGAGATGT ACATCTCAGC AGGAGAGCAC GTCAAGGCCA	130
	TGGAGATCTG TGGTGACCAT GGCTGGGTG ACATGTTGAT CGACATCGCC CCCAACTGG	240
5	ACAAGGCTGA GCGCGAGCCC CTGCTGCTGT GCGCTACCTA CCTCAAGAAG CTGGACAGCC	300
	CTGGCTATGC TGCTGAGACC TACCTGAAGA TGGGTGACCT CAAGTCCCTG GTGCAGCTGC	360
10	AGTGGAGACC CAGCGCTGGG ATGAGGCCCTT TGCTTTGGCT GAGAAGCATC CTGAGTTTAA	420
	GGATGACATC TACATGCCGT ATGCTCAGTG GCTAGCAGAG AACGATCGCT TTGAGGAAGC	480
	CCAGAAAGCG TTCCACAAGG CTGGGCGACA GAGAGAAGCG GTCCAGGTGC TGCAGCAGCT	540
15	CACAAACAAT GCGGTGGCGG AGAGCAGGTT TAATGATGCT GCCTATTATT ACTGGATGCT	600
	GTCCATGCAG TGCTCGATA TAGCTCAAGA TCCTGCCGAG AAGGACACAA TGCTTGCCAA	660
20	GTCTTACCAC TTCCAGCGTT TGGCAGAGCT GTACCATGGT TACCATGCCA TCCATCGCCA	720
	CACGGAAGAT CCGTTCAGTG TCCATCGTCC TGAAACTCTT TTCAACATCT CCAGGTCCT	780
	GCTGCACAGC CTGCCCAAGG ACACCCCTC GGCATCTCT AAAGTCAAAA TACTCTTCAC	840
25	CTTGGCCAAG CAGAGCAAGG CCCTCGGTGC CTACAGGCTG GCCCGGCACG CCTATGACAA	900
	GCTGCGTGGC CTGTACATCC CTGCCAGATT CCAAAAGTCC ATTGAGCTGG GTACCCCTGAC	960
30	CATCCGCGCC AAGCCCTTCC ACGACAGTGA GGAGTTGGTG CCCTGTGCT ACCGCTGCTC	1020
	CACCAACAAC CCGCTGCTCA ACAACCTGGG CAACGTCTGC ATCAACTGCC GCCAGCCCTT	1080
	CATCTTCTCC GCCTCTTCT ACGACGTGCT ACACCTGGTT GAGTTCTACC TGGAGGAAGG	1140
35	GATCACTGAT GAAGAAGCCA TCTCCCTCAT CGACCTGGAG GTGCTGAGAC CCAAGCGGA	1200
	TGACAGACAG CTAGAGATT GCAAACAACA GCTCCAGAT TCTTGGGCT AGTGGGAGAC	1260
40	CAAGGGACTC CATCGGAGAT NAGGACCCGT TCACAGCTAA GCTTAGCTTT GAGCAAGGTG	1320
	GCTCAGATT CGTGCCAGTG GTGTGAGCC GGCTGGTGCT GCGCTCCATG AGCCGCCGGG	1380
	ATGTCTCTAT CAAGCGATGG CCCCCACCCC TGAGGTGGCA ATACTTCCGC TCACTGCTGC	1440
45	CTGACGCTC CATACCATG TGCCCTCTCT GCTTCCAGAT GTTCCATTCT GAGGACTATG	1500
	AGTTGCTGGT GCTTCAGCAT GGCTGCTGCC CTTACTGCCG CAGGTCCAAG GATGACCTG	1560
50	GCCCATGACC AGCATCCTGG GGACGGCCTG CACCTCTGC CCGCCTTGGG GTCTGCTGGG	1620
	CTGTGAAGGA GAATAAGAG TTAAGCTGTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1680
55	AAAAAAAAAA AAAAAAAAAA AANA	1705

(2) INFORMATION FOR SEQ ID NO: 217:

(1) SEQUENCE CHARACTERISTICS:

470

- (A) LENGTH: 999 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

	AGCAATCAC CTTAACGATC TGGAAATGAAA CTGTGACCAG TGCCGCCCTG GGTGTTCTG	60
10	GAGAGACTGC CGTCTTCTTG TTTGCCATA GGTGCTGGG CCCCCGCTC AGTCACTGTC	120
	TCAGACAGKA GTCCCGATAA GCAGATCACC AGTCCTCCAC TGTCTTCCT GTCCGCCCTG	180
	CTGCATGAGA AGATAGCTGC TTCTCCCTC TTTCTCTACA CTGTAAATTA TTGTTTACA	240
15	ATTGAGTGYC TTAATAATAG TYTACAAATA CTATGTATTT ATGCAAACT GTTAAAGTTC	300
	TCATCTGTTA TGATTGGATA CTTGGTCTTG TCAGTAGTGG TCAGCAATGG GTTGTGAGCT	360
20	TGTCTACTC CATACTGTT TATCCTGCTA TGCATTTTAC ATTGTGTGTT CACATCTATT	420
	CCAAGGAGCC TTGCTAGAAA CAACACTGGC GGTTCCTGCA GGCAGGCAG GCATTGCCCC	480
	ATGCTGTGTC CCATAGGAGC CAATGGAAAG AACGTAGCTT GGTCTGCTAG CCAGCCGTGG	540
25	GGTGGCGCAG GCCAGGCAGC CTCTGCACCA GAGTCCAGCA CCTGCCCATT CCCCAGTCAC	600
	ACAATCATACT TCTTCTTTCA TAGAGATTTT ATTACCACCT AGACCACCCT AGTTTCTCTC	660
30	TCTGTTAGTG TCCTGAGCTC TTTTGCAACA AAATGTAGGT ACAGACCAAT CCCTGTCCCT	720
	TCCCCAATCA GGAGCTCCAC ACCATGAGTT GTTTGGTTTT CCAGAAGCTG CCAGTGGGTT	780
	CCCGTGAATT GCGTTAAGAT ATCGATGATK TTTTTATTG TTTTCTTCT TGTTTTTTTA	840
35	AATAATATAT TTAAGGCCAG TATCTTTTGT ACTGTGAATT TGCAGTAGAA GATGCAGAAT	900
	GCACTTTTTT TTTACTTCTG TTGGTGTGTA TTGTATATAG TGTGTGTGCT TCTGTGTGATG	960
40	AAAATAAACT TTTCTTTTAT AAAAAAAAAA AAAAAAAC	999

45

(2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 941 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

55	GGCAGGAGTA GCATTTCATT TAATCTGCAG GTATATTCTC CCAACAGTTT ATTGTGATGT	60
	GATGTCTCTA GCCAAGATTG TPAGGCAGAG AGGAGCTGTC CCAACCTACT ATACCACCGA	120
60	GGCTGGAGAG ATCATATTTT TGGTATTAAA CTGGAGTCTC TCCATCCTTC ACATTGTGTA	180

471

TGTCTCTGT AGCAAACCGG AAAAGTCAGT GACAGAAGAT GCGGCTAGCG GTTTGAGCCA 240
 GAGAATGACA GCTCTGGTTT GGAGAAAAGG CCGGATGGT GCTCTAGAA AGCCCATCTT 300
 5 TCTGCTCTTC TTTTTCCTCC CCGTTATATT GTGCTTTCAT TCATTGATTC ATTGATCAA 360
 CATTTGTTGA GCACCTATTA TGTGTCAAGC TGTGTGCTAG CCTCTGGAAA ACCTGCCCTC 420
 10 ATGTAGCTCA CTGTGGAGTA GGAGAAACAA TGACTACACT ATGATAAGCA CCGGTTGTCA 480
 GGGTCTCACA GAGCAGTGGC CCGTCATCCA GACCGATGAG GTCAAAGAAG GCATCCAGGC 540
 GAGGATGGTG TCAGAGCTAA CTGAAGAATG AGACGGAGCT GCACCAACAG GGGTTGGAAC 600
 15 TGAAGTGGC ACTGCCCTGA GTCTTGATTC CAGCAGAGGG AGACGAGTCT GTGAAAAGGC 660
 ACCAAGGGTG GGAGAGGGCA GAGCAGATGG AGGAAGTTCA GGTAGTTCTG GATGGCCTG 720
 GGGCAAAGCT AGAGAGGTAA GAAGAATCTA CAAATGTTCC TCGAGTTACA TGAAGTTCCA 780
 20 TCCCAATAAA CCCATTGGAA AGGAAAAATT TAAGTCAGAA GTGCTTTTAA GGTGCTCCG 840
 ACTAGAATGA TTTTACAAAC GAATTGATCA CAACCACTTA CAGATGTCTT TGTTCCTTCT 900
 25 CCACTCCAC TGCTTCACCT GACTAGCCTT TAAAAAAA A 941

30 (2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 575 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

40 TAAGTGAAT CCCCCGGGT TGCAGGAAT TGGCAGGAG GAATTCGTAG AAGCTTAAGA 60
 CATACTTTGA AGACAACCTT AGGACCTCC AGCTGCTGG GCATGACCTA CCTTTGCACC 120
 45 CCGCAGTGGT GAAGCCCCAC CTGGGCCATG TTCTGACTA CCTGGTTGCT CCTGCTCTCC 180
 GTGGCCTGGT RCGCCCTCAC AAGAAGCGA AGAAGCTGTC TTCTCTTGT AGGAAGGCCA 240
 AGAGAGCAA GTCCAGAAC CCACTGCGCA GTTCAAGCA CAAAGGAAG AAATTCAGAC 300
 50 CCACAGCAA GCGCTCTGA GGTGTTGGG CCTCTCTGA GGTGAGCACA TTGTGGAGCA 360
 CAGGCTTACA CCTTCTGTG ACAGGCGAGG CTCTGGTGT TACTGCACAG CTTGAACAGA 420
 CAGTTCTGGG GCTGGCAGTG CTGGGCCCTT TAGCTCCTTG GCACTTCCAA GCTGGCATCT 480
 55 TCCCCCTTGA CAACAGATA AAAATTTTAG CTGCCCCAAA AAAAAAAAAA AAAAAAAAAA 540
 CTCGAGGGGG GGGCCGTACC CAATTCGCCC TATAA 575

60

(2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3018 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

	GCCAGCCTTA CAGGTTTAC GTGAAATGAA AGCCATTGGA ATAGAACCCT CGCTTGCAAC	60
15	ATATCACCAT ATTATTGGCC TGTTTGATCA ACCTGGAGAC CTTTAAAGA GATCATCCTT	120
	CATCATTTAT GATATAATGA ATGAATTAAT GGGAAAGAGA TTTTCTCCAA AGGACCCGGA	180
20	TGATGATAAG TTTTTCAGT CAGCCATGAG CATATGCTCA TCTCTCAGAG ATCTAGAACT	240
	TGCTTACCAA GTACATGGCC TTTTAAAAAC CGGAGACAAC TGGAAATTCA TTGGACCTGA	300
	TCAACATCGT AATTTCTATT ATTCCAAGTT CTTCGATTG ATTGTCTAA TGAACAAAT	360
25	TGATGTTACC TTGAAGTGGT ATGAGGACCT GATACCTTCA GCCTACTTTC CCCACTOCCA	420
	AACAATGATA CATCTCTCC AAGCATGGA TGTGGCCAAT CGGCTAGAAG TGATTCCTAA	480
30	AATTTGGGAA AGATAGTAAA GAATATGGTC ATACTTTCCG CAGTGACCTG AGAGAAGAGA	540
	TCCTGATGCT CATGGCAAGG GACAAGCACC CACCAGAGCT TCAGGTGGCA TTTGCTGACT	600
	GTGCTGCTGA TATCAAATCT GCGTATGAAA GCCAACCCTAT CAGACAGACT GCTCAGGATT	660
35	GGCCAGCCAC CTCTCTCAAC TGTATAGCTA TCCTCTTTTT AAGGGCTGGG AGAACTCAGG	720
	AAGCCTGGAA AATGTTGGGG CTTTTCAGGA AGCATAATAA GATTCCTAGX AGTGAGTTGC	780
40	TGAATGAGCT TATGGACAGT GCAAAAGTGT CTAACAGCCC TTCCCAGCCC ATTGAAGTAG	840
	TAGAGCTGGC AAGTGCCCTC AGCTTACCTA TTTGTGAGGG CCTCACCAG AGAGTAATGA	900
	GTGATTTTGC AATCAACCAG GAACAAAAGG AAGCCCTAAG TAATCTAACT GCATTGACCA	960
45	GTGACAGTGA TACTGACAGC AGCAGTGACA GCGACAGTGA CACCAGTGAA GGCAAATGAA	1020
	AGTGGAGATT CAGGAGCAGC AATGGTCTCA CCATAGCTGC TGGAAATCACA CCTGAGAACT	1080
50	GAGATATACC AATATTTAAC ATTGTTACAA AGAAGAAAAG ATACAGATTT GGTGAATTG	1140
	TTACTGTGAG GTACAGTCAG TACACAGCTG ACTTATGTAG ATTTAAGCTG CTAATATGCT	1200
	ACTTAACCAT CTATTAATGC ACCATTAAAG GCTTAGCATT TAAGTAGCAA CATTCGGTT	1260
55	TTCAGACACA TGGTGAGGTC CATGGCTCTT GTCATCAGGA TAAGCCTGCA CACCTAGAGT	1320
	GTGGTGAGC TGACCTCAG ATGCTGTCTT CGTGGGATTG CCCTCTCCTG CTGCTGGACT	1380
60	TCTGCCCTTG TTGGCCTGAT GTGCTGCTGT GATGCTGGTC CTTCACTTA GGTGTTGATG	1440

	CAGTTCTAAC ACAGTTGGGG TTGGGTCAAT AGTTTCCCAA TTTCAGGATA TTTCGATGTC	1500
	AGAAATACCG CATCTTAGGA ATGACTAAAC AAGATAATGG CAGTTTAGGC TGCACAACCTG	1560
5	GTAAAATGAC TGTAGATRAA TGTGTGAATT AGTGTACACG TTGTATTATT TGTTAATATA	1620
	GGCGCTGCCA TAGTTTCTTA ACTTGAACAG CCATGAATGT TTCATGTCTC CCTTTTTTTT	1680
10	TTGTCTATAG CTGTTACCTA TTTTAGTGGT TGAATGAGA GCTAGTGATG ACAGAAGGAT	1740
	GTGGAATGTC TTCTTGACAT CATTTGTGTAT TGCTGGTAAT CAAGTTGGTA ACGACTACTT	1800
	CTAGCAGCTC TTACCACTAT GACTTAAGTG GTCTGGAAG GCAGTAAGTG GAGGTTTGCA	1860
15	GCATTCTGTC CTTCATGAGG GCTTCTACCA CTGACCACTT TGCACGTACC TGGCTCCCAG	1920
	ATTTACTTAG GTACCCACG AGTCGTCCAC ATAGGCAGCT TCATCTTTAC CTTGCCAGAG	1980
20	TTGACAATTA TGGGATACTC TAGTCTACTT ATACTTGTGT TCCCATCTGT CTGCCATCTT	2040
	CTGAAGGCCA GGACCCAGTC ATACATCCTT AGAAACCAAA GTATGGTTTT TGTTTTCTCT	2100
	TGGAATGTCA GGTCTTAAGG CATTTAATTG AGGGACAAA AAAAAAAAAA GCTGATATAG	2160
25	TAGCTAGCTA CTTAAGCATC CATGGGTATT GCTCCATATC AAAGCAGATT TGCAGGACAG	2220
	AAAGAGTAAA TTAGCCTTCA GTCTTGGTTT ACAGCTTCCA AGGAGAGCCT TGGSCACCTG	2280
30	AAATGTTAAC TGGGTCCCTT CCTGTCTCTA GTTCATCAGC ACCTGCAGAT GCTGACTCT	2340
	TGTTAGCCTT ACTATTCAAT ACAGTCTTGA GATTCACGGT ATGCTCTCTC CTATCCAGGC	2400
	ACCTATTCTG AATCACCATG TTGCTCTGCA GCTAGAGTTG ATAGGAGAAA ATCCATTTGG	2460
35	GTAGATGGCC TATGAATTG TAGTAGACTT TCAAAATGAG TGATTTGTTA GCTTGGTACT	2520
	TTTAAGTTTG TGGTACAGAT CCTQCAAACC CATACTCTGA GCAATTAACT GCCTTGAACA	2580
40	TAGAGAAAAA TTAAGGCCTC ACAGGATGAG TCTCCATTCT CTGTAAATGC TTATTTTATC	2640
	ATAGTCTTTA GCCTCTAACT ATGAGTAAAA TGTCTCTCTC GGCCGGGTGT GGTGACTCAC	2700
	ACCTGTAACC TCAGCACTTT GGGAGGCAGA GGTGGGAGGA TCACTTAGGT CCAGGAGTTC	2760
45	GAGACTAGCC TGGGCAACAT AGTGAGACAC CGGATCTACA AAAAAATAA AAGCCAGACT	2820
	GGTGGTATGT ATCTGTGTCC CAGCTAATTG GGAGGGTGAG ATGGGAGGAT TGTTTGAGCC	2880
50	TAGGAGAGCG AGGTTCCAGT GAGCCGTGAT CGCACCCTG CACTCCAGCC TGGGCAACAG	2940
	AGCAAGACCC TGTCTTGAG AAACCAGAAT TTTGGAAGAG CAAATGGGGC TGAGTGCACT	3000
	GGCTCATGCC TGTAAATCC	3018

55

(2) INFORMATION FOR SEQ ID NO: 221:

60

(1) SEQUENCE CHARACTERISTICS:

474.

- (A) LENGTH: 963 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

	GGCAGGAGGG CCGCGGGACA TCCACGGGGC GCGAGTGACA CCGGGGAGGG AGAGCAGTGT	60
10	TCTGCTGGAG CCGATGCCAA AAACCATGCA TTTCTTATTC AGATTGATTG TTTTCTTTTA	120
	TCTGTGGGGC CTTTTTACTG CTCAGAGACA AAAGAAAGAG GAGAGCACCG AAGAAGTGAA	180
15	AATACAAGTT TTGCATCGTC CAGAAAACG CTCTAAGACA AGCAAGAAGG GAGACCTACT	240
	NAAATGCCCC TTATGACGGC TACCTGGCTA AAGACGGCTC GAAATTCTAC TGCAGCCCGA	300
	CACAAAATGA AGGCCACCCC AAATGGTTTG TTCTTGGTGT TGGGCAAGTC ATAAAAGGCC	360
20	TAGACATTGC TATGACAGAT ATGTGCCCTG CAGAAAAGCG AAAAGTAGTT ATACCCCTTT	420
	CATTTCATA CCGAAAGGAA GGCTATGCAG AAGGCAAGAT TCCACCGGAT GCTACATTGA	480
25	TTTTTGAGAT TGAACTTTAT GCTGTGACCA AAGGACCACG GAGCATTGAG ACATTTAAAC	540
	AAATAGACAT GGACAATGAC AGGCAGCTCT CTAAGCCGA GATAAACCTC TACTTGCAAA	600
	GGGAATTTGA AAAAGATGAG AAGCCACGTG ACAAAGTCATA TCAGGATGCA GTTTTAGAAG	660
30	ATATTTTAA GAAGAATGAC CATGATGGTG ATGGCTTCAT TTCTCCCAAG GAATACAATG	720
	TATACCAACA CGATGAACATA TAGCATATTT GTATTTCTAC TTTTTTTTTT TAGCTATTTA	780
35	CTGTACTTTA TGTATWAAAC AAAGTCMCTT TTCTCCMAGT TGAATTTGCT ATTTTTCGCC	840
	TATGAGAAGA TATTTTGATC TCCCAATAC ATTGATTTTG GTATAATAAA TGTGAGGCTG	900
	TTTTGCAAC TTAATAAAAA ATTTAAAAAA ACTGGAGGGG GGCCCGTACC CAATTCGCCG	960
40	NATATGAT	963

45

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1404 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

55	CGTTTTCCGG CCGTGGGTTT GTGGCCGTCC GGCCTCCCTG ACATGCAGCC CTCTGGACCC	60
	CGAGGTTGGA CCTACTGTG ACACACCTAC CATGCCGACA CTCTTCAACC TCCTCTGGGT	120
60	TGCCCTGGCC TGCAGCCCTG TTCACACTAC CTTGTCAAAG TCAGATGCCA AAAAAGCCGC	130

	CTCAAAGACG CTGCTGGAGA AGAGTCAGTT TTCAGATAAG CCGGTGCAAG ACCGGGGTTT	240
	GGTGGTGACG GACCTCAAAG CTGAGAGTGT GGTTCCTTGG CATCCAGCTT ACTGCTCCGC	300
5	AAAGGCCCCG GACAGACACT TTGCTGGGGA TGTACTGGGC TATGTCACCTC CATGGAACAG	360
	CCATGGCTAC GATGTCACCA AGGTCTTTGG GAGCAAGTTC ACACAGATCT CACCCGTCTG	420
10	GCTGCAGCTG AAGAGACGTG GCCGTGAGAT GTTTGAGGTC ACGGGCCTCC ACGACGTGGA	480
	CCAGGGGTGG ATGCGAGCTG TCAGGAAGCA TGCCAAGGGC CTGCACATAG TGCCTCGGCT	540
	CCTGTTTGAG GACTGGACTT ACGATGATTT CCGGAACGTC TTAGACAGTG AGGATGAGAT	600
15	AGAGGAGCTG AGCAAGACCG TGGTCCAGGT GGCAAGAAC CAGCATTTCTG ATGGCTTCGT	660
	GGTGGAGGTC TGAACCCAGC TGCTAAGCCA GAAGCGCGTG GGCCTCATCC ACATGCTCAC	720
20	CCACTTGGCC GAGGCTCTGC ACCAGGCCCG GCTGCTGGCC CTCCTGGTCA TCCCGCCTGC	780
	CATCACCCCC GGGACCGACC AGCTGGCCAT GTTCACGCAC AAGGAGTTTG AGCAGCTGGC	840
	CCCCGTGCTG GATGGTTTCA GCTCATGAC CTAGGACTAC TCTAGAGCGC ATCAGCCTGG	900
25	CCCTAATGCA CCCCTGTCTT GGGTTCGAGC CTGGCTCCAG GTCTTGAGC CGAAGTCCAA	960
	GTGGCGAAGC AAAATCCTCC TGGGGCTCAA CTTCTATGGT ATGGACTACG CGACCTCCAA	1020
30	GGATGCCCCG GAGCCTGTTG TCGGGCCAG GTACATCCAG AACTGAAGG ACCACAGGCC	1080
	CCGGATGGTG TGGACAGCC AGGCTCAGA GCACTTCTTC GAGTACAAGA AGAGCCGCAG	1140
	TGGGAGGCAC GTCGTCTTCT ACCCAACCTT GAAGTCCCTG CAGGTGCGGC TGGAGCTGGC	1200
35	CCGGGAGCTG GCCGTTGGGG TCTCTATCTG GGAGCTGGCC AGGGCCTGGA CTACTTCTAC	1260
	GACCTGCTCT AGGTGGGCAT TGGGGCCTCC GCGGTGGACG TGTTCTTTTC TAAGCCATGG	1320
40	ACTGAGTGAG CAGGTGTGAA ATACAGCCCT NCACTCCGTT TGCTGTGAAA AAAAAAAAAA	1380
	AAAAAAAAA AAAAAAAAAA AAAA	1404

45

(2) INFORMATION FOR SEQ ID NO: 223:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 707 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

50

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

55

	NGCGCGCCTG CAGTCGACAC TAGTGGATCC AAAGAATTCG GCACGAGGGC AGGTCCAGGG	60
	CTCAGAAATC AGCTCTATTG ACGAATTCTG CCGCAAGTTC CGCCTGGACT GCGCGCTGGC	120
60	CATGGAGCCG ATCAAGGAGG ACCGCCCCAT CACCATCAAG GACGACAAGG GCAACCTCAA	180

CCGCTGCATC GCAGACGTGG TCTGGCTTTT GAGCAGCGTC ATGACACATC TGGGCTTCCA 240
 GATCGCGCGC ATGGATGAGA TCGAGGCTCA CTGCGGAGAG CTGACGTAAT GATGACACCG 300
 CATGAGCGAC CTCCACCGCG ACTTTGAGGG CGGCGAGAGG GTGAGCTAT GTTTCAGAC 360
 CCGAGCGCGC ATGTGCGGCT CAGATGAGCT GACACATCA CAGGTGCTC AGATGCTGTT 420
 CGACCTGGAG TCAGGCTACA ACGGCTTCCA CGGCTTCTG CATGCTTAC CCGGCGGAC 480
 TAGCCCTTGC ACAGAAGGGC AGATGCTGAG GCGATGCTC CTGCTGCTT GTGCGGACA 540
 CAGGCGGCTG TCATCGACAC AACTGCTGCT CTGAGCTGCT CTGCTGCTG TGTGCTTGT 600
 GTGTGGAAC TTGTGGGCGG GCGGCTGCGG CAGATGAG ATGCTGCTG ACCTTCAAAA 660
 AAAAAAAAAA AAAAAGCTCG GGGGGGCGG GTGCGATCC CCGGCTT 707

(2) INFORMATION FOR SEQ ID NO: 224:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1384 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

GGGGAAGTGC AGTGACAGCA GCGTAAAGAG TGGGAGGCGG GACAGAGTTC GACACAGGT 60
 ATGGAGAGGG GGTTCAGCGA GCCTAGAGAG GCGAGATAT CAGGCTGCTG GCGGTGAGAA 120
 TCCAGGGAGA GGAGCGGAAA CAGAGAGGGG GAGAGAGACC GGGGAGTTC TGGGTTGCG 180
 AGCCGCTCAG CCGTGTTCGG AGCGAGGCGA CAGTGGCTAC CAGGTCTTCT ACAGGTCCC 240
 GGGCTGCGCT TGGTCTGCTT GCTTCTGCTC CTGCGGCGCG GGTGCGGCTCA GAGGGGTCA 300
 GAGCCGCTCC TGCTGGAGGG GAGTGGCTTG GTGCTGCTG AGGCTGCTCG AGTGTCTCA 360
 GGGGGGCGCG GGGGAGGCGC CTGCGGAGAG GCGAGGCTTG GCGAGTGGC ACTTCTGCG 420
 GTCCGAAGCC AAGCAGATCA GCGAGAGGG GAGAGGCGA ATGCGATTC TGGGCGATC 480
 TACTTCGACC AGGTCTGCTT GAAGAGGGG GGTGGCTTTG ACCGGGCTTC TGGTCTCTC 540
 GTAGCCCTTG TCCGGGCTGT CTACAGCTTC CGGTTCTATG TGTGAGCTT GTACAACCG 600
 CAACTGTCC AGGTGAGCTT GATGCTGAG ACCTGCGCTG TCATCTGAGC CTGTCGCAAT 660
 GATCTGAGC TGACCGGGGA GCGAGGAGC AGCTCTGCTT TACTGCTCTT GCGGCTGCG 720
 GACCGAGTGT CTCTGCGGCT GCGTGGGGG AATCTGCTG GTGCTTGA AACTCAACT 780
 TTCTCTGCTT TCTCATCTT CCGTCTCTG GAGCGAGT TTGCTGAGCA CAGATGCA 840

477

5
10
15
20

GGCCCTGACA ACTTTCTTCT GGCCTCTCTT GCGCCAGAAA CAGCAGAGGC AGGAGAGAGA 900
CTCCCTCTGG YTCCTATCCC ACYTCTTTGC ATGGGAMCCT GTGCCAAACA CCCAAGTTTA 960
AGAAAARARY ARARCTGAGG CAGGTATACA GAGCTGGAAG TGGACCATGG AAAACATSGA 1020
TAACCATGCA TC/TCTTCT TGGCCACCTC CTGAAACTGT CCACCTTTGA AGTTTGAAC 1080
TTAGTCCCTC CAACTCTGA CTGCTGCCCTC CTTCCTCCCA GCTCTCTCAC TGAGTTATYT 1140
TCACTGTACC TGTTCAGCA TATCCCCACT ATCTCTCTTT CTCCTGATCT GTGCTGTCTT 1200
ATTCTCTCC TTAGGCTTCC TATTACCTGG GATTCATGA TTCATTCTT CAGACCTCT 1260
CCTGCCAGTA TGCTAAACCC TCCCTCTCTC TTCTTATCC CGCTGTCCCA TTGGCCACG 1320
CTGGATGAAT CTATCAATAA AACAACCTAGA GAATGGTGGT CAAAAAAAAA AAAAAAAAAA 1380
TCGA 1384

(2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 760 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

35
40
45
50
55
60

GGGTCCACCC ACGCGTCCGC TGACCACTCC GTTATAGATA CTTCTTCTTA TACCAAAACT 60
GTTTAAACAG GTGCCACCAC AAGGGATGTC GTCCTTACTC TCTGCGGGTC TTCAAGCATC 120
CCTTGTGGG AAAGTCTCT GGGCAAGCAC GTGGTATTTG GTCTGCTGCT TGCTTCCCTT 180
TTTCCACCAG GGATGTTGTG ATCATAAGTC AAAACAACAG TATATTCCAA ATCTCAAAAG 240
CTATGTGGC CTGAGCACAA TTGAAATCTA GCAGAGTTT TCCTATGTAG CTTTAGAGTA 300
ACTCTTCTGC TTCTCTGTCA CTTACAATTC AGGTTCTGCC TTTGCCTAAG AGCATGAGCA 360
GAAGAGTCCT CATGTGACGC TTAGTTCTAT TGCACTCCTG GGTGAAACTA TTAAAGCAT 420
GGGGTGCTK CTCCCCAMWT CCTCCCTAAC AATTCGTTGT GTGGACTTCT CATCTAAAAG 480
GTTAGTGGCT TTTGCTTGGG ATCAGTGCTC TGTATTGATG TTCTTGCTGG TCTCCAGACA 540
CATTCCTGTT GCATTAAGAC TTGAAAGACT TGTAGATGTG TGATGTTTCA GCACAGGATG 600
CTGAAAGCTA TGTTACTATT CTTAGTTTGT AAMTGTCTT TTTGATACCA TCATCTGTT 660
TTCTTTTGT AGGTATAAAT AAAAACAATG TTGACAATAA AAAAAAAAAA AAAAAAAAAA 720
AAAAAAAAA AAAAAAAAAA NAAAAAAAAA AAAAAAAAAA 760

(2) INFORMATION FOR SEQ ID NO: 226:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2057 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

CCGAGCCGGC TCGCGCGGGG GAATCCGTGC GGGCGCTTC CGTCCCGTTC CCATCCTCGC 60
CGCGCTCCAG CACCTCTGAA GTTTTCGAGC GCCCAGAAAG GAGGCGAGCA AGGAGGGAGT 120
GTGTGAGAGG AGGGAGCAAA AAGCTCAGCC TAAACATTT ATTTCAAGGA GAAAAGAAAA 180
AGGGGGGGCG CAAAAATGGC TGGGGCAATT ATAGAAAACA TGAGCACCAA GAAGCTGTGC 240
ATTGTGTGGT CGATTCTGCT CGTGTTCOA ATCATCGCCT TTCTGGTGGG AGGCTTGATT 300
GCTCCAGGGC CCACAACGGC AGTGTCTAC ATGTCCGTGA AATGTGTGGA TGCCCGTAAG 360
AACCATCACA AGACAAAATG GTTCGTGCTT TGGGGACCCA ATCATTGTGA CAAGATCCGA 420
GACATTGAAG AGGCAATTCC AAGGGAAATT GAAGCCAATG ACATCGTGT TTCTGTTCAC 480
ATTCCTCTCC CCCACATGGA GATGAGTCTT TGGTTCCAAT TCATGTTGTT TATCCTGCAG 540
CTGGACATTG CCTTCAAGCT AAACAACCAA ATCAGRGAAA ATGCAGAAGT CTCCATGGAC 600
GTTTCCCTGG CTTACCGTGA TGACCGGTTT GCTGAGTGA CTGAAATGGC CCATGAAAGA 660
GTACCACGGA AACTCAAATG CACCTTCACA TCTCCCAAGA CTCCAGAGCA TGGAGGGCCG 720
GTTACTATGA ATGTGATGTC CTTCCTTCA TGGAAATTGG GTCTGTGGCC CATGAAGTTT 780
TACCTTTTAA ACATCCCGCT GCCTGTGAAT GAGAAGAAGA AAATCAATGT GGGAAATGGG 840
GAGATAAAGG ATATCCCGTT GGTGGGGATC CACCAAAATG GAGCCTTCAC CAGGTGTGG 900
TTTGCCATGA AGACCTTCTT TACGCCCAGC ATCTTCATCA TTATGGTGTG GTATTGGAGG 960
AGGATCACCA TGATGTCCCG ACCCCCAGTG CTTCTGGAAG AAGTCATCTT TGCCCTTGGG 1020
ATTTCCATGA CCTTTATCAA TATCCCAGTG GAATGGTTTT CCATCGGTTT TGAATGGACC 1080
TGGATGCTGC TGTGTGTGA CATCCGACAG GCATCTTCTA TGCRATGCTT CTCTCCTTCT 1140
GGATCATCTT CTGTGGCGAG CACATGATG ATCAGCACGA GCGGAACCA ATGCGAGGGT 1200
ATTGGAACCA AGTCGGACCC ATTGCGGTTG GTCTTCTGCT CTCTTCATAT TTGACATGTG 1260
TGAGAGAGGG GTACAACCTA CGAATCCCTT CTACAGTATC TGGACTACAG ACATTTGGGA 1320
CAGAGCTGGC CATGGCTTTC ATCATGTTGG CTGGAATCTG CCTCTGCCTC TAACTTCCTG 1380
TTCTATGCT TCATGGTATT TCAGGTGTTT CGGAACATCA GTGGGAAGCA GTCCAGCCTG 1440

	CGAGTATGTA GGAATGCGG TGGGTGAGG GATGAGGGGG TAAATTTTATG GTTCAGATTG	1500
	CTCAGGTGTA TCACTTGGG TTGGGTGGG ATGATGTTCA TTTTCTTCAT CGTTAGTCAG	1560
5	GTAAGGGAAG GGAATGCGG AATGGGGGGG GGTGAGGTTG GCAAGTGAAC AGTCCCTTTT	1620
	TGCAAGGAT GTTGGGAGG TGGATGTTT AGTCTTTTGG TGTGATGTTG TTGTATGCAC	1680
10	CATCCCATTA AACTATGTA GAGAGCAAT GCAAGGAT GCACTCCCA TGTAAATCGA	1740
	GGAAGATTG TGTCTGTTT GTTGGGAGG TTTATGAGA ACTGTTGAGG GCTTCGAAAT	1800
	ATTCTTTTCA CATGAGCAAT TGAATTTTG TATTGTAAT GACCAAGGCA ACACATGTTT	1860
15	ATCAGTTTGG CATTTGCAAT TTTGAGGTTG AATTTGATTG TACTTGTATA CGCACACAAA	1920
	TACACTGAT TACCTTTTAT GTGAAATTT TAAATATAG GAAAAAAGCG TCAACAATAA	1980
	ATATCTTTG AATATGTTT TACTCTCTT AAAAAAAA AAAAAAATG GTGCCGAATT	2040
20	CGGACGAGG GGTACCA	2057

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2084 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

35	GGGAGGAGG CATTTCTTGG AAGAGGCGA ACCCGCATTC CTCTGTGCCC CTCTCTCCCC	60
	ACCAATGCTT TTTGAAATAT AGTCTCTTTC ACCGGAATA ACTGTTGATT TTTCACCTCT	120
40	CCCTCTTAGG TCACACTTTT GAGAAAAA ATGTGCAAGC TGGAAACCAG AAGAAAAATA	180
	TGAGACGGGG AATGACCTTG TAACTGTGTT SGTGCTTTTG GTTGAGTGTG TGGAGTCCTG	240
45	CTCAGGTGTT AATTAGAGTG TGTGTGATCG TGGTGGTTTG AAGCGAACCG CTGTTCAGA	300
	GCTGTGACTG CCGCTGCAAT GAGAGAGAC TGGCTTTGGG TGCTCGTAGC GCGGGGCTTT	360
	CTCTCTTGGT CATCTGCAAG AGCGGCAAT GTCCGGGAGG CAGAAGGTAC CGGGCCAGCT	420
50	ACTGGAGGAG TGTGGGGGGC TGGCTGGGTT GGGGCTGCG CCGTGGGGCC CTGTTGCTGC	480
	TGTCCATGTA TTTCTATGAC TTTCTGCGA ATGCTGTTGG CCGGGCTTTC ACTTGGATGC	540
55	TGGGCTTCTT GGGGCTTCTG GAGGGCAATG AATATCTTCC TGGGCTTCAA GGGGCTGGCC	600
	CGAGCTGAGA TGTGTGCAAT GTGTGAAAA GGGATTTTCA ACGTGGCCCA TGGGCTGGCA	660
	TGCTCATATT AATCTGGAAT TGTGGGGTGG ATGTGCGGAG AGCTCGAGGC CCGGATTGGA	720
60	ACTTACGATC AATATGAGA GACCTGCTTA CGGGTGGAG TGAGCCAGCG GTGTATATAT	780

480

CTCTCCGCT TGGACTGTGG GGTGCTGAT AACCTGAGTA TGGCTGACCC CAACATTGGC 840
 TTCCTGGATA AACTGCCCCA GCAGACGGT GACCTGCTG GCATCAAGGA TCGGGTTTAC 900
 5 AGCAACAGCA TGTATGAGCT TCTGGAGAAC GGGCAGCGGG CGGGCACCTG TGTCTGGAG 960
 TACGCCACCC CTTGCGAGAC TTGTTTGGC ATGTCACAAT ACAGTCAAGC TGGCTTTAGC 1020
 10 GGGGAGGATA GCTTGAGCA GGCCAACTC TTCTGCCGA CACTGAGCA CATCTGGCA 1080
 GATGCCCCCTG ACTCTCAGAA CAACTGCCGC CTCATTGCCCT ACCAGGAACC TGCAGATGAC 1140
 AGCAGCTTCT CGCTGTCCA GGAGTTTCT CGGCACCTGC GGCAGGAGGA AAAGGAAGAG 1200
 15 GTTACTGTGG GCAGCTTGAA GACCTCAGCG GTGCCAGTA CCTCCAGCAT GTCCCAAGAG 1260
 CCTGAGCTCC TCATCAGTGG AATGGAAAAG CCCCTCCCTC TCCGCACCGA TTCTCTTGA 1320
 20 GACCCAGGCT CACCAGGCCA GAGCCTCCAG TGGTCTCCAA GCCTCTGGAC TGGGGGCTCT 1380
 CTTCACTGGC TGAATGTCCA GCAGAGCTAT TTCCTTCCAC AGGGGGCTT GCAGGGAAGG 1440
 GTCCAGGACT TGACATCTTA AGATGCGTCT TGTCCCTTG GCCCAGTCAT TTCCCTCTC 1500
 25 TGAGCCTCGG TGTCTTCAAC CTGTGAAATG GGATCATAAT CACTGCCCTA CCTCCCTCAC 1560
 GGTGTGTGTG AGGACTGACT GTGTGGAAGT TTTTCATAAA CTTTGGATGC TAGTGTACTT 1620
 30 AGGGGGTGTG CCAGGTGTCT TTGATGGGGC CTTCCAGACC CACTCCCCAC CTTCTCCGC 1680
 TTCCTTTGCC CGGGGACGCC GAACTCTCTC AATGGTATCA ACAGGCTCCT TCGCCCTCTG 1740
 GCTCCTGGTC ATGTTCCATT ATTGGGGAGC CCCAGCAGAA GAATGGAGAG GAGCAGGAGG 1800
 35 CTGAGTTTGG GGTATTGAAT CCCCCGGCTC CCACCCTGCA GCATCAAGGT TGCTATGGAC 1860
 TCTCCTGCCG GGCAACTCTT GCGTAATCAT GACTATCTCT AGGATTCTGG CACCACTTCC 1920
 40 TTCCCTGGCC CCTTAAGCCT AGCTGTGTAT CGGCACCCCC ACCCCACTAG AGTACTCCCT 1980
 CTCACTTGGG GTTTCCTTAT ACTCCACCCC TTTCTCAACG GTCCTTTTTT AAAGCACATC 2040
 45 TCAGATTAAA AAAAAAAAAA AAAAAAAAAA AGGGGGGGCN GCTT 2084

(2) INFORMATION FOR SEQ ID NO: 228:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2143 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

TGCACCGACG CGTCCGGTTG AATTCCTTGA CTTGCAACA CATATTTATT AGCTGACTC

60

	AAACAATGAA GCTATTAAAA CTTCGGAGGA ACATTGTAAA ACTCTCTTTG TATCGGCATT	120
	TCACCAACAC GCTTATTTTG GCAGTGGCAG CATCCATTGT GTTTATCATC TGGACAACCA	130
5	TGAAGTTCAG AATAGTGACA TGTCAGTCGG ACTGGCGGGA GCTGTGGGTA GACGATGCCA	240
	TCTGGCGCTT GCTGTTCTCC ATGATCCTCT TTGTATCAT 'GGTCTCTGG CGACCATCTG	300
10	CAACCAACCA GAGGTTTGCC TTTTCACCAT TGTCTGAGGA AGAGGAGGAG GATGAACAA	360
	AGGAGCCTAT GCTGAAAGAA AGCTTTGAAG GAATGAAAAT GAGAAGTACC AAACAAGAAC	420
	CCAATGGAAA TAGTAAAGTT AACAAAGCAC AGGAAGATGA TTTGAAGTGG CTAGAAGAGA	480
15	ATGTTCCCTT TCTGTGACA GATGTAGCAC TTCCAGCCCT TCTGCATTCA GATGAGGAAC	540
	GAATGATCAC ACACTTTGAA AGGTCCAAAA TGGAGTAACG AATGGGAAGA TTTGCAGTTA	600
20	AAGATGGCTA CCATCAGGGA AGAGATCAGC ATCTGTGTCA GTCTTCTGTA CGGCTCCATG	660
	CGATTAAAGG AACCAATGAC ATCCTGATCT GTTCCCTTGAT CTTTGGGCAT TGGAGTTGGC	720
	GAGAGGTGTC AGAACAAGA GAACATCTTA CTGAAAACAA GTTCATAAGA TGAGAAAAAT	780
25	CTACGAGCTT CTTATTTACA AACTGCTGC CCCCTTTCTT CCCAGACTCT GACATGGATG	840
	TTCAATGCAAC TTAAGTGTGT TGTTCCTGAA CTTTCTGTAA TGTTCATTT TTTAAATCTG	900
30	ACAACTAAA AAGTTTAACG TCTTCTAAAA GATTGTATC AACACCATAA TATGTAATCT	960
	CCAGGAGCAA CTGCCTGTAA TTTTATTTA TTTAGGGAGT TACATAGGTG ATGGGGGAAA	1020
	TGTTTAACTA CTTTTCATTT TCCTGGGAAG TCAAGGTTAC ATCTTGCAGA GGTGTGTTTG	1080
35	AGAAAAAAGG GCCCTTCTGA GTTAAGGAGC CATAGTTCTA TCAATGATCA AAAGAAAAA	1140
	AAAAAAAGA GAACTGTTA CAGTATGATT CAGATCATTT AAAAAAGCAA AATCAAGTGC	1200
40	AATTTTGTIT ACAATGGTG TATATTAAAG ATTTTCTAT TTCAGATGTA CTTTAAAGAG	1260
	AAATATTAGC TTAATCTTT TGACATCTGC TATGTGACA CATCCCATG CTGGCAATGT	1320
	GGTGCACACT CCGAAACTTT TAACTACTGT TTTGTAAGCC TCCAAGGCTG GCATTGCAGG	1380
45	GTCCCTAGGC AATGTTTTGT TTGCCCTTAT GCAGAGAGGT GCTCCAAGTG CTGTGATTGA	1440
	GCACCCTGCT AGAGGAACTG TAATGCTTCA GAAGTTGTAG CTTATACAAA GGAAACAGGT	1500
50	CCTGCTGGCT TAATTTAAAC AGTTATTGCA TGAAGTAGCG TGGAGGCCCT GGACTGCTGC	1560
	TGTTCTTTA GGATGGACTG TTCTGGTATC TGGTATTGGT TTAGAGACTG TTAATAAGGG	1620
	ACATCACAAG GTGATGGGAT TCATTTGAAG CACTCTATTT CTGTTTAAAT GGTTTATCC	1680
55	AATTTTGCCT TCCCAAGATT TTTGTTCTAC ATAAAAAGTT CATGCCACTT TTTAATATAA	1740
	AAAAATTTAA CAAAATTAAT GTATTTTCT CATTTTTC AAACTTTTC TAAAGACTCT	1800
60	TTCTGTCAAA CTCATGAAAA ATTTCTTTCT ATGGCTTTTA TTCTAGATTG TCTTATTTTC	1860

482

TGTTAAAACC AATGACCACA TGACCACAAT CTTCACTAAC TCATACTGCA GTGAAAGTGT 1920
TAACCCCTTAG GTAGTTTCTC TACAACCTCT TGCTATGGTG ATTTTAAAAA AAGTTTCCTA 1980
5 GCGAAGTATC TCTGAGGGA CAGGCAATCT GAAGGAACTG ACTATATTCT CCATGGCTAA 2040
GTCCATTAGG CCAAAAGNCT GGGTGGGTAT TGGTTGTGCA GCTGTCTATT GGCATATTAA 2100
AAACGTAGGC CGGANGGAAT AATTAGGTTG TATGCGCGGC GGG 2143

(2) INFORMATION FOR SEQ ID NO: 229:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1025 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

CCTGCCCCAC ATTGCTTCAT TGGCCTGGCC ATGCGCCTCT ACTATGGCAG CCGCTAGTCC 60
25 CTGACAACTT CCACCTGAT TCCGGACCCT GTAGATTGGG CGCCACCACC AGATCCCCCT 120
CCCAGGCCCTT CCTCCCTCTC CCATCAGCAG CCGTGTAAAC AGTGCCTTGT GAGAAAAGCT 180
30 GGAGAAGTGA GGCAGCCAG GTTATTCTCT GGAGGTTGGT GSATGAAGGG GTACCCTAGG 240
AGATGTGAAG TGTGGGTTTG GTTAAGGAAA TGCTTACCAT CCCCCACCCC CAACCAAGTT 300
CTTCCAGACT AAAGAATTAA GGTAAACATCA ATACCTAGGC CTGAGAAATA ACCCCATCCT 360
35 TGTGGGCAG CTCCCTGCTT TGTCTGCTAT GAACAGAGTT GATGAAAGTG GGGTGTGGGC 420
AACAAGTGGC TTTCCTTGGC TACTTTAGTC ACCCAGCAGA GCGACTGGAG CTGGCTAGTC 480
40 CAGCCCAGCC ATGCTGCTAT ACTCTTCCAT AAGGGATCCT CACCCCTCCA CTTTCATGCA 540
AGAAGGCCCA GTTGCCACAG ATTATAACAAC CATTACCCAA ACCACTCTGA CAGTCTCCTC 600
CAGTTCCAGC AATGCCTAGA GACATGCTCC CTGCCCCCTC CACAGTGCTG CTCCCCACAC 660
45 CTAGCCTTTG TTCTGGAAAC CCCAGAGAGG GCTGGGCTTG ACTCATCTCA GGAATGTAG 720
CCCCTGGGCC CTGGCTTAAG CCGACACTCC TGACCTCTCT GTTCACCCCTG AGGGCTGTCT 780
50 TGAAGCCCGC TACCCACTCT GAGGCTCCTA GGAGGTACCA TGCTTCCCAC TCTGGGGCCT 840
GCCCCCTGCT AGCAGTCTCC CAGCTCCCAA CAGCCTGGGG AAGCTCTGCA CAGAGTGACC 900
TGAGACCAGG TACAGGAAAC CTGTAGCTCA ATCAGTGTCT CTTTAACTGC ATAAGCAATA 960
55 AGATCTTAAT AAAGTCTTCT AGGCTGTAGG GTGGTTCCCTA CAACCACAGC CAAAAAATA 1020
AAAAA 1025

(2) INFORMATION FOR SEQ ID NO: 230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1250 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

```

5      G C C C A C G C G T C C G C C C A C G C G T C C G G G C G G T G C G G A G T A T G G G C G C T G A T G G C C A T G G A G      60
15     G C C T A C T G C C G C T T C C T G G C G C Y G C T G G G G T C G G C A C T G C T C G T C G G C T T C C T G T C G G T G      120
      A T S T T C G C C C T C G T C T G G G T C C T C C A C T A C C G A G A G G G G C T T G G C T G G G A T G G G A G C G C A      130
20     C T A G A G T T T A A C T G G C A C C C A G T G C T S A T G G T C A C C G G C T T C G T C T T C A T C C A G G G C A T C      240
      G C A T C A T C G T C T A C A C A C T G C C G T G G A C C T G G A A A T G C A G C A A G C T C C T G A T G A A A T C C A      300
      T C C A T G C A G G G T T A A A T G C A G T T G C T G C C A T T C T T G C A A T T A T C T C T G T G T G G C C C G T G T      360
25     T T G A G A A C C A C A A T G T T A A C A A T A T A G C C A A T A T G T A C A G T C T G C A C A G C T G G G T T G G A C      420
      T G A T A G C T G T C A T A G C T A T T G T T A C A G C T T C T T C A G G T T T T T C A G T C T T C T G C T T C      480
      C A T G G G C T C C G C T T C T C T C C G A G C A T T T C T C A T G C C C A T A C A T G T T T A T T C T G G A A T T G      540
30     T C A T C T T T G G A A C A G T G A T T G C A C A G C A C T T A T G G G A T T G A C A G A G A A A C T G A T T T T T T      600
      C C C T G A G A G A T C C T G C A T A C A G T A C A T T C C C G C C A A G G T G T T T T C G T A A A T A C G C T T G      660
35     G C C T T C T G A T C C T G G T G T T C G G G C C C T C A T T T T T G G A T A G T C A C C A G A C C G C A A T G G A      720
      A A C G T C C T A A G G A G C C A A A T T C T A C C A T T C T T C A T C C A A A T G G A G G C A C T G A C A G G G A G      780
      C A A G A G G T T C C A T G C C A G C C T A C T C T G G C A A C A A C A T G G A C A A A T C A G A T T C A G A G T T A A      840
40     A C A R T G A A G T A G C A G C A A G G A A A A G A A A C T A G C T C T G G A T G A G G C T G G G C A G A C A T C T A      900
      C C A T G T A A A A T G T T G T A G A G A T A G A G C C A T A T A A C G T C A C G T T T C A A A A C T A G C T C T A C A      960
45     G T T T T G C T T C T C C T A T T A G C C A T A T G A T A A T T G G C C T A T G T A G T A T C A A T A T T A C T T T A      1020
      A T C A C A A A G G A T G G T T T C T T G A A A T A A T T T G T A T T G A T T G A G C C T A T G A A C T G A C C T G A      1080
      A T T G G A A A G G A T G T G A T T A A T A A A T A A T A G C A G A T A T A A A T T G T G G T T A T G T T A C C T T      1140
50     T A T C T T G T T G A G G A C C A C A C A T T A G C A C G G T G C C T T G T G C A K A A T A G A T A C T C A A T A T G      1200
      T G A A T A T G T G T C T A C T A G T A G T T A A T T G G A T A A A C T G G C A G C A T C C C T G A      1250

```

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:

484

- (A) LENGTH: 1811 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

	CNENCAGTAC CGGTGNGATT CCGGGGTGGA CCCACGGCTC CGCTGCATTG CAGGGGCTTT	60
10	CAGTGGCTTT CATTTCTGAAG TTCCTGGATA ACATGTTCCA TGTCTTGATG GCCCAGGTTA	120
	CCASTGTGAT TATCACAACA GTGTCTGTCC TGGTCTTTGA CTTCAGGCCC TCCTTGAAT	180
	TTTTCTTGA AGCCSCATCA GTCSTYCTCT CTATATTTAT TTATAATGCC AGCAAGCCTC	240
15	AAGTTCCGGA ATACGCACCT AGGCAAGAAA GGATCCGAGA TCTAAGTGGC AATCTTTGGG	300
	AGCGTTCCAG TGGGGATGGA GAAGAACTAG AAAGACTTAC CAAACCCAAG AGTGATGAGT	360
20	CAGATGAAGA TACTTTGTAA CTGGTACCCA CATAGTTTGC AGCTCTCTTG AACCTTATTT	420
	TCACATTTTC AGTGTTTGTA ATATTTATCT TTTCACTTTG ATAAACCAGA AATGTTTCTA	480
	AATCCTAATA TTCTTTGCAAT ATATCTAGCT ACTCCCTAAA TGGTTCCATC CAAGGCTTAG	540
25	AGTACCCAAA GGCTAAGAAA TTCTAAAGAA CTGATACAGG AGTAACAATA TGAAGAATTC	600
	ATTAATATCT CAGTACTTGA TAAATCAGAA AGTTATATGT CCAGATTATT TTCCTTGGCC	660
30	TTCAAGCTTC CAAAAAATT GTAATAATCA TGTTAGCTAT AGCTTGTATA TACACATAGA	720
	GATCAATTTG CCAAATATTC ACAATCATGT AGTTCTAGTT TACATGCCAA AGTCTTCCCT	780
	TTTTAACATT ATAAAAGCTA GGTGTCTCT TGAATTTTGA GGCCCTAGAG ATAGTCATTT	840
35	TGCAAGTAA GAGCAACGGG ACCCTTTCTA AAAACGTTGG TTGAAGGACC TAAATACCTG	900
	GCCATACCAT AGATTTGGGA TGATGTAGTC TGTGCTAAAT ATTTTGCTGA AGAAGCAGTT	960
40	TCTCAGACAC AACATCTCAG AATTTTAATT TTAGAAATT CATGGGAAT TGGATTTTGT	1020
	TAATAATCTT TTGATGTTTT AAACATTGGT TCCCTAGTCA CCATAGTTAC CACTTGTATT	1080
	TTAAGTCATT TAAACAAGCC ACGGTGGGGC TTTTCTCTCC TCAGTTTGAG GAGAAAAATC	1140
45	TTGATGTGAT TACTCCTGAA TTATTACATT TTGGAGAATA AGAGGGCATT TTATTTTATT	1200
	AGTTACTAAT TCAAGCTGTG ACTATTGTAT ATCTTTCCAA GAGTTGAAAT GCTGGCTTCA	1260
50	GAATCATACC AGATTGTGAG TGAAGCTGAT GCCTAGGAAC TTTTAAAGGG ATCCTTTCAA	1320
	AAGGATCACT TAGCAACAC ATGTTGACTT TTAAGTATG TATGAATATT AATACTCTAA	1380
	AAATAGAAAG ACCAGTAAATA TATAAGTCAC TTTACAGTGC TACTTCACAC TTAAAAGTGC	1440
55	ATGGTATTTT TCATGGTATT TTGCATGCAG CCAGTTAACT CTCGTAGATA GAGAAGTCAG	1500
	GTGATAGATG ATATTAAAAA TTACCAACA AAAGTGACTT GCTCAGGGTC ATGCAGCTGG	1560
60	GTGATGATAG AAGAGTGGGC TTAACTGGC AGGCCTGTAT GTTTACAGAC TACCATACTG	1620

TAAATATGAG CTTTATGGTG TCATTCTCAG AAACCTATAC ATTTCTGCTC TCCTTTCTCC 1630
 TAAGTTTCAT GCAGATGAAT ATAAGGTAAT ATACTATTAT ATAATTCAAT TGTGATATCC 1740
 ACAATAATAT GACTGGCAAG AATGGGTGGA AATTTGTAAT TAAAATTAAT ATTAAACCTA 1800
 AAAAAAAAAA N 1911

(2) INFORMATION FOR SEQ ID NO: 232:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2371 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CTGACCTCAT GCGGTAGAGC CTAGCAACAG CGCAGGCTCC CAGCCGAGTC CGTTATGGCC 60
 GCTGCCGTCC CGAAGAGGAT GAGGGGGCCA GCACAAGCGA AACTGCTGCC CGGGTCGGCC 120
 ATCCAAGCCC TTGTGGGGTT GCGCGGGCCG CTGGTCTTGG CGCTCTGCT TGTGTCCGCC 180
 GCTCTATCCA GTGTTGTATC ACGGACTGAT TCACCGAGCC CAACCGTACT CAACTCACAT 240
 ATTTCTACCC CAAATGTGAA TGCTTTAACA CATGAAAACC AAACCAAACC TTCTATTGCC 300
 CAAATCAGCA CCACCTCCC TCCACGACG AGTACCAAGA AAAGTGGAGG AGCATCTGTG 360
 GTCCCTCATC CCTCGCTAC TCCTCTGTCT CAAGAGGAAG CTGATAACAA TGAAGATCCT 420
 AGTATAGAGG AGGAGGATCT TCTGATGCTG AACAGTTCTC CATCCACAGC CAAAGACACT 480
 CTAGACAATG GCGATTATGG AGAACCAGAC TATGACTGGA CCACGGGCCC CAGGGACGAC 540
 GACGAGTCTG ATNGACACCT TCGAAGAAAA CAGGGGTTAC ATGGAAATTG AACAGTCAGT 600
 GAAATCTTTT AAGATGCCAT CCTCAATAT AGAAGAGGAA GACAGCCATT TCTTTTTTCA 660
 TCTTATTATT TTTGCTTTTT GCATTGCTGT TGTTTACATT ACATATCACA ACAAAGGAA 720
 GATTTTCTT CTGGTTCAAA GCAGGAAATG GCGTGATGGC CTTTGTCCA AACAGTGA 780
 ATACCATCGC CTAGATCAGA ATGTTAATGA GGCAATGCCT TCTTTGAAGA TTACCAATGA 840
 TTATATTTTT TAAAGCACTG TGATTTGAAT TTGCTTATGT AATTTTATTT GCTTGACTTT 900
 TTATATGATA TTGTGCAAT GTTTGCCATA GGCAATTGGT ACTTAAATGA GAGGTGAGTC 960
 TCTCTTTTGC CTGGTGCTT TCGAAATTA ATGTCACAAA CGAGTATATA ATTTTTTATC 1020
 TGTACTTTTA GAGCTGAGTT TAATCAGTG TCCAAATGT GAGTTAAACA TTACCTTATA 1080
 TTTACACTGT TAGTTTTTAT TGTTTTAGAT TTATTATGCT TCTTCTGGAA GTATTAGTGA 1140

486

TGCTACTTTT AAAAGATCCC AAACCTGTAA CTAATTTCTG ACATATCTGT TACTGCTGAC 1200
 TCACATTCAT TCTCCGCCAT TCAATACTA TTTTTATCC ACATTTTTTT TTGTTCCCA 1260
 5 ACTGTAATGT ACAAGGATAT GTGTGATAAT GCTTTGGATT TGAGTAATAT TTTTTTTCT 1320
 TCCAAGAAAA CTGCTTTGGA TATTTTAGA TAATTTAAAC ATAATTTAGG ATAATGATAT 1380
 10 TGCTCAATCT GACCACAATT TTAGGTAAAA CATTAAATGT GTCAAGAAAT CTGCGCAACA 1440
 GAGACTCTGC AGCTTGCACT GGACATAGAT AAAATGTTAC AGAGATACTA TTTTTTGGT 1500
 TGGAACTACT ATATTAAATT TAGAAGCAGA AACTGGTAAA ATGTTAAATA CATGTACAAT 1560
 15 TGCTTTTAGT TAGCAATTGA TTGTAGCATG GGTTCCTCCA AGGTTTCAAG CAATGGGCAG 1620
 AGTTTAAAT TATATCAGAT TCGTTTACTT CGTTTATTAT TTTACAGTAA ATTGAATAA 1680
 ATCTTAGGGG TCATTATCAC TTAAATAATA CTGTACCTAG GTCTTTCAAA TTAAATTAT 1740
 20 ACCTGAATGA AGTTGTTTGT ATACATAAAG GATATTTGTG TACAATTACC TTTTTCCTCC 1800
 CACACTTGTG TTCTTTGTTT TTGTTTTTA TGGCAACTGG AAAGTATTTA CTATGGGATT 1860
 25 CATTTATGTC TGTCTTCTA TCATAAAGAA TTGATCAATA TGTAATATG TGATTTGAAC 1920
 CATGGTTGAC TTACAAGTGT CACTACAGCT TTTTAGAAAA CATAGCCCTA ATATATGTTA 1980
 AGCAGGACCC GGGTGACCA GTGGGCTTGC GCTTTATGTA GAGCTGGAAG AAGGCCGTCC 2040
 30 ATCCTGTCTC TTGGGCGGAC AGTGTACTTT CCTAATAGGG AAGGGAAGCA CAATGGAAAT 2100
 ACCCTGAAC CGTTTTATTG CAGTAATTTT TTTCATATCT GAAACTATTA TTAAATATT 2160
 35 TGAATAAGAT TTTAAAAAAT AAATGGCAA GATATAAATC TAAAAAANA AAAAAAANA 2220
 AAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA N 2271
 40

(2) INFORMATION FOR SEQ ID NO: 233:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1338 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

50 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

CTTCCGGTTC TCCGGGCAGC TGCCACTGCT GTAGCTTCTG CCACCTGCCA CGACCGGGCC 60
 TCTCCCTGGC GTTTGGTCAC CTCTGCTTCA TTCTCCACCG CGCCTATGGT CCTCTTTGGA 120
 55 GCCAGCGTGG CGNGCCTGGC GGCTCCCGGG TGGTGAGAGA GCGGTCCGGG AACGATGAAG 180
 GCCTGGCAGT GCTGCTGCTG TCTCAGCCAC CTCTTGGCTT CCGTCTCTCT CTTGCTGTG 240
 60 CTGCCTGAAC TAAGCGGGYC CCTGGMAGTC CTGCTGCAGG CAGCGGAGGC CCGGCCAGGT 300

YTTGGGGCTC CTGACCCTAG ACCAGGACAT TACCGCCGCT GCCACGGGGC COTWACCCCT 360
 GCCCAGGCAGC CCGGCCGTGG TCTGGGTGAA GCTGCGGGGG CCGCGGGGCT CCGAGGGAGG 420
 5 CAATGGCAGC AACCCCTGTGG CCGGGCTTGA GACGGACGAT CACGGAGGGA AGCCCGGGGA 480
 ARGCTCGGTG GGTGGCGGCC TTGCTGTGAG CCCCACCCCT GCGGACAAGC CCATGACCCA 540
 10 GCGGGCCCTG ACCGTGTTGA TGGTGGTGAG CCGCGCGGTG CTGGTGTACT TCGTGGTCAG 600
 GACGGTCAGG ATGAGAAGAA GAAACCGAAA GACTAGGAGA TATGGAGTTT TGGACACTAA 660
 CATAGAAAAT ATGGAATTGA CACCTTTAGA ACAGGATGAT GAGGATGATG ACAACACGTT 720
 15 GTTTGATGCC AATCATCCTC GAAGATAAGA ATGTGCCCTT TGATGAAAGA ACTTTATCTT 780
 TCTACAATGA AGAGTGAAT TTCTATGTTT AAGGAATAAG AAGCCACTAT ATCAATGTTG 840
 20 GGGGGGTATT TAACTTACAT ATATTNAAC AACCTTTAAT TTGCTGTTGC AATAAATACC 900
 GTATCCTTTT ATTATATCTT TATATGTATA GAAGTACTCT GTTAATGCGC TCAGAGATGT 960
 TGGGGATAAA GTATACTGTA ATAATTTATC TGTTTGAAAA TTAATAATAA ACGGTGTTTT 1020
 25 CTGRTCGGTT TTTGTTTCCT GCTTACCATA TGATTGTAAA TTGTTTTATG TATTAATCAG 1080
 TTAATGCTAA TTATTTTTCG TGATGTGATA TGTTAAAGAG CTATAAATTC CAACAACCRA 1140
 30 CTGGTGTGTA AAAATAATTT AAAATYTCTT TTAAGTAAAG GTATTTCCCA TTTTGTGGG 1200
 GAAAAGAAGC CAAATTTATT ACTTTGTGTT GGGGTTTTTA AAATATTAAG AAATGTCTAA 1260
 GTTATTGTTT GCAAAACAAT AAATATGATT TTAAATTCTC TTAAAAAAA AAAAAAAC 1320
 35 CCGGGGGGGG GGCCCGGN 1338

40

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

Met Leu Ser Thr Gly Ile Glu Val Ala Arg Pro Pro Ala Thr Leu Leu
 1 5 10 15

Gly Leu Met Phe Val Leu Thr Gly Met Pro Arg Gly Leu Arg Xaa
 20 25 30

55

(2) INFORMATION FOR SEQ ID NO: 235:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 116 amino acids

60

488

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

5 Met Asn Val Val Ile Val Ile Ile Leu Phe Ser Phe Asp Ser Val Gly
 1 5 10 15
 Thr Met Phe Ser Cys Asn Arg Ile Pro Lys Ile Thr Val Leu Asn Lys
 20 25 30
 10 Leu Lys Phe Xaa Cys Glu Val Leu Leu Arg Ile Gln Thr Ile Gln Gly
 35 40 45
 Phe Tyr Arg Cys Thr Arg Ile Ser Arg Tyr Lys Gly Ile Phe Pro Asp
 50 55 60
 Phe Cys Gln Ser Gln Cys Met Gly Cys Asn Pro Glu Ser Xaa Met Ala
 65 70 75 80
 20 Val Pro Ala Leu Val Thr Pro Ile Leu Ala His Arg Lys Lys Glu Lys
 85 90 95
 Gly Met Cys Leu Phe Thr Leu Ile Ile Ala Pro Thr Arg Cys Thr His
 100 105 110
 25 Tyr Phe Cys Xaa
 115

30

(2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 103 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

40 Met Ser Ser Ala Lys Ile Val Arg Gln Arg Gly Ala Val Pro Thr Tyr
 1 5 10 15
 Tyr Thr Thr Glu Ala Gly Glu Ile Ile Phe Leu Val Leu Asn Trp Ser
 20 25 30
 45 Leu Ser Ile Leu His Ile Val Asp Val Leu Cys Ser Lys Pro Glu Lys
 35 40 45
 Ser Val Thr Glu Asp Ala Ala Ser Gly Leu Ser Gln Arg Met Thr Ala
 50 55 60
 50 Leu Val Trp Arg Lys Gly Pro Asp Gly Gly Ser Arg Lys Pro Ile Leu
 65 70 75 80
 Leu Leu Phe Phe Phe Leu Pro Leu Ile Leu Cys Phe His Ser Phe Ile
 85 90 95
 55 His Ser Ser Asn Ile Cys Xaa
 100

60

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

Met Ile Leu Phe Pro Gln Xaa Ala Leu Arg Leu Gly Xaa Trp Pro Arg
 1 5 10 15
 Thr Trp Ser Ile Leu Xaa Lys Tyr Ser Val Asn Phe Phe Ser Ala Tyr
 20 25 30
 Ser Pro Met Gly Ala Val Gly Thr Glu Phe
 35 40

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

Met Ile Ile Leu Leu Leu Phe Met Leu Leu Asn Asn Val Val Leu Val
 1 5 10 15
 Gln Glu Asp Asn Cys Gln Arg Lys Asn Thr Val Gln Glu Arg Arg Xaa
 20 25 30
 Trp Ser Gln Trp Xaa
 35

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 128 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

Met Ala Ala Xaa Pro Pro Gly Cys Thr Pro Pro Xaa Leu Leu Asp Ile
 1 5 10 15
 Ser Trp Leu Thr Glu Ser Leu Gly Ala Gly Gln Pro Val Pro Val Glu
 20 25 30
 Cys Arg His Arg Leu Glu Val Ala Gly Pro Arg Lys Gly Pro Leu Ser
 35 40 45
 Pro Ala Trp Met Pro Ala Tyr Ala Cys Gln Arg Pro Thr Pro Leu Thr
 50 55 60
 His His Asn Thr Gly Leu Ser Glu Leu Leu Glu His Gly Val Cys Glu

65					70						75					80
Glu	Val	Glu	Arg	Val	Arg	Arg	Ser	Glu	Arg	Tyr	Gln	Thr	Met	Lys	Val	
				85					90					95		
Arg	Arg	Ala	Gly	Leu	Gly	Pro	Thr	Pro	Gly	Met	Ser	Cys	Pro	Gly	Asn	
			100					105					110			
Asp	Asn	Thr	Val	His	Thr	Met	His	Gly	Glu	Ala	Asn	Arg	Gly	Ser	Xaa	
			115				120					125				

Met	Ser	Thr	Phe	Gln	Leu	Leu	Leu	Ile	Leu	Ala	Gln	Ser	Thr	Tyr	
1				5				10						15	
Lys	Ile	Lys	Ser	Lys	Pro	Leu	His	Met	Thr	Asn	His	Thr	Leu	Leu	Asn
			20					25						30	
Ser	Pro	Gly	Leu	Asn	Pro	Ser	Ser	Pro	Thr	Leu	Asn	Phe	Lys	Thr	Gln
		35					40					45			
Gln	His	Glu	Ser	Val	Ser	Tyr	Ala	Cys	Cys	His	Met	Arg	Ser	Leu	His
50						55					60				

His Ala Phe Ala Xaa
65

5

(2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

Met Val Ser Val Val Leu Ile Phe Ser Phe Leu Ser Leu Thr Ile Ser
1 5 10 15
Thr Thr Ala Ser Ala Tyr Asn Gly Asn Asp Thr Gln Gly Trp Asn Asp
20 25 30
Lys Phe His Xaa Xaa Ser Val Lys Thr Gln Thr Xaa
35 40

20

25

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

Met Ile Ser Asp Ala Gly Ala Gly Phe Gly Val Phe Leu Leu Val Pro
1 5 10 15
Arg Ala Gly His Cys Trp Gly Ala Gly Lys Pro Leu Pro Ser Cys Pro
20 25 30
Ser Val Ala Ser Ile Pro Ser Trp Val Leu Pro Ser Phe Leu Glu Arg
35 40 45
Gly Arg Xaa
50

45

(2) INFORMATION FOR SEQ ID NO: 244:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

Met Val Gln Thr Ile Gln Asp Phe Leu Ser Leu Phe Ser Thr Pro Ile
1 5 10 15
Phe Leu Leu Leu Leu Met Phe Glu Thr Leu Ser Leu Ala Pro Ala Trp
20 25 30

60

492

Leu Lys Pro Leu Arg Val Thr Ser His Ser Xaa
 35 40

5.

(2) INFORMATION FOR SEQ ID NO: 245:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 61 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

Met Ile Leu Met Pro Gly Leu Gly Thr Ser Arg Gln Arg Ser Val Pro
 1 5 10 15
 Phe Val Pro Thr Leu Asn Ala Ser Thr Pro Gly Ala Met Thr Gly Pro
 20 25 30
 Thr Ala Thr Leu Thr Ser Cys Gln Trp Thr Thr Ala Cys Arg Val Ser
 35 40 45
 Trp Ala Asn Gly Trp Thr Ser Leu Arg Thr Phe Arg Xaa
 50 55 60

30

(2) INFORMATION FOR SEQ ID NO: 246:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 36 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

Met Ser His His Ala Gln Pro Arg Phe Leu Leu Ile Thr Met Leu Leu
 1 5 10 15
 Gln Glu Ala Lys Pro Val Ser Asn Ile Pro His Leu Leu Glu Ser Trp
 20 25 30
 Tyr Phe Gly Xaa
 35

45

(2) INFORMATION FOR SEQ ID NO: 247:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

Met Asn Ser Leu Phe Trp Met Ile Leu Leu Pro Val Ser Gln Asp Gln
 1 5 10 15
 Val Val Glu Gly Leu Gln Gly Gly Phe Ser Gln Ile His Met Arg Ile
 20 25 30

60

493

Leu Arg Lys His Leu Xaa
35

5

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 211 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

15

Met Ser Arg Ser Xaa Asp Val Thr Asn Thr Thr Phe Leu Leu Met Ala
1 5 10 15

Ala Ser Ile Tyr Leu His Asp Gln Asn Pro Asp Ala Ala Leu Arg Ala
20 25 30

20

Leu His Gln Gly Asp Ser Leu Glu Cys Thr Ala Met Thr Val Gln Ile
35 40 45

25

Leu Leu Lys Leu Asp Arg Leu Asp Leu Ala Arg Lys Glu Leu Lys Arg
50 55 60

Met Gln Asp Leu Asp Glu Asp Ala Thr Leu Thr Gln Leu Ala Thr Ala
65 70 75 80

30

Trp Val Ser Leu Ala Thr Gly Gly Glu Lys Leu Gln Asp Ala Tyr Tyr
85 90 95

Ile Phe Gln Glu Met Ala Asp Lys Cys Ser Pro Thr Leu Leu Leu Leu
100 105 110

35

Asn Gly Gln Ala Ala Cys His Met Ala Gln Gly Arg Trp Glu Ala Ala
115 120 125

40

Glu Gly Leu Leu Gln Glu Ala Leu Asp Lys Asp Ser Gly Tyr Pro Glu
130 135 140

Thr Leu Val Asn Leu Ile Val Leu Ser Gln His Leu Gly Lys Pro Pro
145 150 155 160

45

Glu Val Thr Asn Arg Tyr Leu Ser Gln Leu Lys Asp Ala His Arg Ser
165 170 175

His Pro Phe Ile Lys Glu Tyr Gln Ala Lys Glu Asn Asp Phe Asp Arg
180 185 190

50

Leu Val Leu Gln Tyr Ala Pro Ser Ala Glu Ala Gly Pro Glu Leu Ser
195 200 205

55

Gly Pro Xaa
210

(2) INFORMATION FOR SEQ ID NO: 249:

60

494

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 548 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

5 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

```

Met Glu Asp Ser Glu Ala Leu Gly Phe Glu His Met Gly Leu Asp Pro
 1           5           10           15

10 Arg Leu Leu Gln Ala Val Thr Asp Leu Gly Trp Ser Arg Pro Thr Leu
    20           25           30

Ile Gln Glu Lys Ala Ile Pro Leu Ala Leu Glu Gly Lys Asp Leu Leu
    35           40           45

15 Ala Arg Ala Arg Thr Gly Ser Gly Lys Thr Ala Ala Tyr Ala Ile Pro
    50           55           60

Met Leu Gln Leu Leu Leu His Arg Lys Ala Thr Gly Pro Val Val Glu
20 63           70           75           80

Gln Ala Val Arg Gly Leu Val Leu Val Pro Thr Lys Glu Leu Ala Arg
    85           90           95

25 Gln Ala Gln Ser Met Ile Gln Gln Leu Ala Thr Tyr Cys Ala Arg Asp
    100          105          110

Val Arg Val Ala Asn Val Ser Ala Ala Glu Asp Ser Val Ser Gln Arg
    115          120          125

30 Ala Val Leu Met Glu Lys Pro Asp Val Val Val Gly Thr Pro Ser Arg
    130          135          140

Ile Leu Ser His Leu Gln Gln Asp Ser Leu Lys Leu Arg Asp Ser Leu
35 145          150          155          160

Glu Leu Leu Val Val Asp Glu Ala Asp Leu Leu Phe Ser Phe Gly Phe
    165          170          175

40 Glu Glu Glu Leu Lys Ser Leu Leu Cys His Leu Pro Arg Ile Tyr Gln
    180          185          190

Ala Phe Leu Met Ser Ala Thr Phe Asn Glu Asp Val Gln Ala Leu Lys
    195          200          205

45 Glu Leu Ile Leu His Asn Pro Val Thr Leu Lys Leu Gln Glu Ser Gln
    210          215          220

Leu Pro Gly Pro Asp Gln Leu Gln Gln Phe Gln Val Val Cys Glu Thr
50 225          230          235          240

Glu Glu Asp Lys Phe Leu Leu Leu Tyr Ala Leu Leu Lys Leu Ser Leu
    245          250          255

55 Ile Arg Gly Lys Ser Leu Leu Phe Val Asn Thr Leu Glu Arg Ser Tyr
    260          265          270

Arg Leu Arg Leu Phe Leu Glu Gln Phe Ser Ile Pro Thr Cys Val Leu
    275          280          285

```

60

495

Asn Gly Glu Leu Pro Leu Arg Ser Arg Cys His Ile Ile Ser Gln Phe
 290 295 300
 5 Asn Gln Gly Phe Tyr Asp Cys Val Ile Ala Thr Asp Ala Glu Val Leu
 305 310 315 320
 Gly Ala Pro Val Lys Gly Lys Arg Arg Gly Arg Gly Pro Lys Gly Asp
 325 330 335
 10 Lys Ala Ser Asp Pro Glu Ala Gly Val Ala Arg Gly Ile Asp Phe His
 340 345 350
 His Val Ser Ala Val Leu Asn Phe Asp Leu Pro Pro Thr Pro Glu Ala
 355 360 365
 15 Tyr Ile His Arg Ala Gly Arg Thr Ala Arg Ala Asn Asn Pro Gly Ile
 370 375 380
 Val Leu Thr Phe Val Leu Pro Thr Glu Gln Phe His Leu Gly Lys Ile
 385 390 395 400
 Glu Glu Leu Leu Ser Gly Glu Asn Arg Gly Pro Ile Leu Leu Pro Tyr
 405 410 415
 25 Gln Phe Arg Met Glu Glu Ile Glu Gly Phe Arg Tyr Arg Cys Arg Asp
 420 425 430
 Ala Met Arg Ser Val Thr Lys Gln Ala Ile Arg Glu Ala Arg Leu Lys
 435 440 445
 30 Glu Ile Lys Glu Glu Leu Leu His Ser Glu Lys Leu Lys Thr Tyr Phe
 450 455 460
 Glu Asp Asn Pro Arg Asp Leu Gln Leu Leu Arg His Asp Leu Pro Leu
 465 470 475 480
 His Pro Ala Val Val Lys Pro His Leu Gly His Val Pro Asp Tyr Leu
 485 490 495
 40 Val Pro Pro Ala Leu Arg Gly Leu Val Arg Pro His Lys Lys Arg Lys
 500 505 510
 Lys Leu Ser Ser Ser Cys Arg Lys Ala Lys Arg Ala Lys Ser Gln Asn
 515 520 525
 45 Pro Leu Arg Ser Phe Lys His Lys Gly Lys Lys Phe Arg Pro Thr Ala
 530 535 540
 50 Lys Pro Ser Xaa
 545

(2) INFORMATION FOR SEQ ID NO: 250:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

Met Thr Thr Val Pro Pro Ser Pro Arg Pro Met Ser Arg Pro Ser Glu
 1 5 10 15
 5 Arg Asn Met Arg Arg Pro Arg Gly Pro Ser Pro Leu Pro Ala Ser Pro
 20 25 30
 Arg Asn Ser Thr Pro Asp Glu Pro Asp Val His Phe Ser Lys Lys Phe
 35 40 45
 10 Leu Asn Val Phe Met Ser Gly Arg Ser Arg Ser Ser Ala Glu Ser
 50 55 60
 Phe Gly Leu Phe Ser Cys Ile Ile Asn Gly Glu Glu Gln Glu Gln Thr
 65 70 75 80
 His Arg Ala Ile Phe Arg Phe Val Pro Arg His Glu Asp Glu Leu Glu
 85 90 95
 20 Leu Glu Val Asp Asp Pro Leu Leu Val Glu Leu Gln Ala Glu Asp Tyr
 100 105 110
 Trp Tyr Glu Ala Tyr Asn Met Arg Thr Gly Ala Arg Gly Val Phe Pro
 115 120 125
 25 Ala Tyr Tyr Ala Ile Glu Val Thr Lys Glu Pro Glu His Met Ala Ala
 130 135 140
 Leu Ala Lys Asn Ser Asp Trp Val Asp Gln Phe Arg Val Lys Phe Leu
 145 150 155 160
 30 Gly Ser Val Gln Val Pro Tyr His Lys Gly Asn Asp Val Leu Cys Ala
 165 170 175
 35 Ala Met Gln Lys Ile Ala Thr Thr Arg Arg Leu Thr Val His Phe Asn
 180 185 190
 Pro Pro Ser Ser Cys Val Leu Glu Ile Ser Val Arg Gly Val Lys Ile
 195 200 205
 40 Gly Val Lys Ala Asp Asp Ser Gln Glu Ala Lys Gly Asn Lys Cys Ser
 210 215 220
 His Phe Phe Gln Leu Lys Asn Ile Ser Phe Cys Gly Tyr His Pro Lys
 225 230 235 240
 Asn Asn Lys Tyr Phe Gly Phe Ile Thr Lys His Pro Ala Asp His Arg
 245 250 255
 50 Phe Ala Cys His Val Phe Val Ser Glu Asp Ser Thr Lys Ala Leu Ala
 260 265 270
 Glu Ser Val Gly Arg Ala Phe Gln Gln Phe Tyr Lys Gln Phe Val Glu
 275 280 285
 55 Tyr Thr Cys Pro Thr Glu Asp Ile Tyr Leu Glu
 290 295
 60

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

Leu Leu Tyr Leu Leu Lys Val Xaa Val Ile Phe Val Phe Ser Ser Ser
 1 5 10 15
 Lys Gly Val Thr Leu Val Ser Met Asn Leu Thr Ser Phe Phe Val Ser
 20 25 30
 Ser Val Leu Ala Cys Phe Ser Xaa
 35 40

(2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 594 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

Met Pro Ala Ser Ser Leu Glu Ser Arg Ser Phe Leu Leu Ala Lys Lys
 1 5 10 15
 Ser Gly Glu Asn Val Ala Lys Phe Ile Ile Asn Ser Tyr Pro Lys Tyr
 20 25 30
 Phe Gln Lys Asp Ile Ala Glu Pro His Ile Pro Cys Leu Met Pro Glu
 35 40 45
 Tyr Phe Glu Pro Gln Ile Lys Asp Ile Ser Glu Ala Ala Leu Lys Glu
 50 55 60
 Arg Ile Glu Leu Arg Lys Val Lys Ala Ser Val Asp Met Phe Asp Gln
 65 70 75 80
 Leu Leu Gln Ala Gly Thr Thr Val Ser Leu Glu Thr Thr Asn Ser Leu
 85 90 95
 Leu Asp Xaa Leu Cys Tyr Tyr Gly Asp Gln Glu Pro Ser Thr Asp Tyr
 100 105 110
 His Phe Gln Gln Thr Gly Gln Ser Glu Ala Leu Glu Glu Glu Asn Asp
 115 120 125
 Glu Thr Ser Arg Arg Lys Ala Gly His Gln Phe Gly Val Thr Trp Arg
 130 135 140
 Ala Lys Asn Asn Ala Glu Arg Ile Phe Ser Leu Met Pro Glu Lys Asn
 145 150 155 160
 Glu His Ser Tyr Cys Thr Met Ile Arg Gly Met Val Lys His Arg Ala
 165 170 175

498

Tyr Glu Gln Ala Leu Asn Leu Tyr Thr Glu Leu Leu Asn Asn Arg Leu
 180 185 190

5 His Ala Asp Val Tyr Thr Phe Asn Ala Leu Ile Glu Ala Thr Val Cys
 195 200 205

Ala Ile Asn Glu Lys Phe Glu Glu Lys Trp Ser Lys Ile Leu Glu Leu
 210 215 220

10 Leu Arg His Met Val Ala Gln Lys Val Lys Pro Asn Leu Gln Thr Phe
 225 230 235 240

Asn Thr Ile Leu Lys Cys Leu Arg Arg Phe His Val Phe Ala Arg Ser
 245 250 255

15 Pro Ala Leu Gln Val Leu Arg Glu Met Lys Ala Ile Gly Ile Glu Pro
 260 265 270

20 Ser Leu Ala Thr Tyr His His Ile Ile Arg Leu Phe Asp Gln Pro Gly
 275 280 285

Asp Pro Leu Lys Arg Ser Ser Phe Ile Ile Tyr Asp Ile Met Asn Glu
 290 295 300

25 Leu Met Gly Lys Arg Phe Ser Pro Lys Asp Pro Asp Asp Asp Lys Phe
 305 310 315 320

Phe Gln Ser Ala Met Ser Ile Cys Ser Ser Leu Arg Asp Leu Glu Leu
 325 330 335

30 Ala Tyr Gln Val His Gly Leu Leu Lys Thr Gly Asp Asn Trp Lys Phe
 340 345 350

35 Ile Gly Pro Asp Gln His Arg Asn Phe Tyr Tyr Ser Lys Phe Phe Asp
 355 360 365

Leu Ile Cys Leu Met Glu Gln Ile Asp Val Thr Leu Lys Trp Tyr Glu
 370 375 380

40 Asp Leu Ile Pro Ser Ala Tyr Phe Pro His Ser Gln Thr Met Ile His
 385 390 395 400

Leu Leu Gln Ala Leu Asp Val Ala Asn Arg Leu Glu Val Ile Pro Lys
 405 410 415

45 Ile Trp Lys Asp Ser Lys Glu Tyr Gly His Thr Phe Arg Ser Asp Leu
 420 425 430

50 Arg Glu Glu Ile Leu Met Leu Met Ala Arg Asp Lys His Pro Pro Glu
 435 440 445

Leu Gln Val Ala Phe Ala Asp Cys Ala Ala Asp Ile Lys Ser Ala Tyr
 450 455 460

55 Glu Ser Gln Pro Ile Arg Gln Thr Ala Gln Asp Trp Pro Ala Thr Ser
 465 470 475 480

Leu Asn Cys Ile Ala Ile Leu Phe Leu Arg Ala Gly Arg Thr Gln Glu
 485 490 495

60